Karlova Univerzita v Praze

Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE

Mgr. Jana Veverková

Statistické odhady a chvosty jejich rozdělení pravděpodobností

Katedra pravděpodobnosti a matematické statistiky

Vedoucí: prof. RNDr. Jana Jurečková, DrSc.

Studijní obor: Pravděpodobnost, matematická statistika a ekonometrie

Studijní plán: Matematická statistika

Akademický rok: 2011/2012

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

Prague, August 4, 2012

I would like to thank my supervisor prof. RNDr. Jana Jurečková, DrSc. for her very kind approach and valuable and professional advices.

Abstract

Master Thesis Statistical estimators and their tail behavior provides description of two type of characteristics of robustness of estimators - tail behavior and breakdown point. Description is made for translation equivariant estimators in general and also for some concrete type of estimators, sample mean, sample median, trimmed mean, Huber estimator and Hodges Lehmann estimator. Tail behavior of these estimator is illustrated for random sample coming from t-distribution with 1 to 5 degrees of freedom. Ilustration is based on simulations made in Mathematica.

Author	Mgr. Jana Veverková
Department	Department of Probability and Mathematical Statistics
Supervisor	prof. RNDr. Jana Jurečková, DrSc.
Keywords	Statistical estimator, light-tailed distribution, heavy-tailed distribution, breakdown point

Abstrakt

Diplomová práce Statistické odhady a chvosty jejich rozdělení pravděpodobností popisuje dva typy charakteristik robustnosti - míru chování chvostů a bod selhání. Popis je zaměřen na odhady ekvivariantní vzhledem k posunutí, nejprve v obecné smyslu a pak také pro konkrétní typy odhadů. Jedná se o průměr, medián, useknutý průměr, Huberův odhad a Hodges Lehmannův odhad. Míra chování chvostů těchto odhadů je ilustrována pro náhodné výběry pocházející z t-rozdělení. Tato ilustrace je provedena na základě simulací vyhotovených v programu Mathematica.

Autor	Mgr. Jana Veverková
Katedra	Katedra pravděpodobnosti a matematické statistiky
Vedoucí	prof. RNDr. Jana Jurečková, DrSc.
Klíčová slova	Statistický odhad, rozdělení s lehkými chvosty, rozdělení s těžkými chvosty, bod selhání odhadu

Contents

Chapter 1

Introduction

For statistical models when we are not sure about the true type of distribution which the random sample comes from, the asymptotic approach is often used. Although the true distribution is not known, many estimators have asymptotically normal distribution under some conditions and this fact is often used especially for testing hypothesis. The true distribution can be heavy tailed or contaminated by heavy tailed one which could cause outliers to appear in our sample. However the asymptotic normal distribution has exponential tails hence the problems of outliers seems to be solved by using asymptotic approach. Nevertheless the asymptotic distribution is true only for infinite sample size and it is obvious that any random sample we make is always finite. Then it is possible to choose rather some robust estimator which is insensitive to changes in the initial distribution and hence also to outliers. Also for the commonly known robust estimators the asymptotic normality holds. However, this is still not the end of the story since it was shown, that for any finite sample size the distribution of translation equivariant estimators, among which the most commonly used robust estimators belong, is heavy tailed whenever the initial distribution is heavy tailed as well.

These problems were discussed by many authors in previous decades. Jurečková [1979] dealt with characteristics of robustness of L-estimators and compared them with a sample mean. Moreover she described the tail behaviour of the distribution of a sample mean when the random sample comes from a heavy tailed distribution. Jurečková [1981] then added the tail behaviour of sample mean when the initial distribution is exponential tailed and discussed also tail behaviour of robust estimators. By He, Jurečková, Koenker and Portnoy [1990] the relationship between the tail behaviour of estimators and another characteristic of robustness, which is a breakdown point, was introduced. This relationship was well established for monotone and translation equivariant estimators. This paper was followed by Kušnier and Mizera [2001]. They dealt with possibility of generalization of previous conclusions on wider family of estimators, especially on translation and scale equivariant, but not necessarily monotone. Many other characteristics and properties of robust estimators were discussed by spme other authors, as asymptotic or finite sample distributions for example.

In this thesis we will describe two characteristics of robustness, the tail behaviour and the breakdown point. We will show their relationship as was shown in named papers. Furthermore, we describe and compare some not only robust estimators according to these characteristics. This will be completed by simulations. Using simulations we will show and compare the tail behaviour of some estimators when the initial sample is from the Student t-distribution. This distribution has some interesting properties, for example it approaches the normal distribution for degrees of freedom going to infinity. However, for any finite number of degrees of freedom it is heavy tailed. When having random sample coming from t-distribution it could tempt to work with approximation by the normal distribution. However, distribution of any translation equivariant estimator based on such sample is heavy tailed. From this point of view it is interesting to look at the tail behaviour of such estimators for different degrees of freedom.

The thesis is divided as follows. Besides first chapter which is introductory there are another four chapters. The second chapter describes tail behaviour and breakdown point generally as characteristics of robustness. The third chapter describes these characteristics of sample mean, M, L and R-estimators. Results of simulations are shown in this chapter. The fourth chapter describes the main idea of how it can be shown that the asymptotic distribution of used estimators is normal but heavy tailed for any finite sample size. The fifth chapter concludes.

Chapter 2

Characteristics of robustness

2.1 Idea of robustness

Let $\mathbb{X} = (X_1, \ldots, X_n)$ be a random sample from a population with a distribution P defined on σ-algebra B. Suppose that P is an element of the family of distributions $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, where $\Theta \subseteq \mathbb{R}^p$, $p \in \mathbb{N}$, is the parametric space.

The most usually we assume that the probability distribution P is known and the only missing information is the value of the parameter θ. In such case our interest is restricted to finding some appropriate estimator of θ and evaluating its properties, moreover, we can be interested in testing hypothesis. If we know the probability distribution P, there are methods developed in statistics for estimating θ such as the method of moments or the method of maximum likelihood for example. As far as the evaluation of used estimator properties is concerned, it might be necessary to have a concrete type of distribution P. The same is true for hypothesis testing. Especially, many methods are useful only under assumption of the normal probability distribution which is the case of Student's t-test or Fisher-Snedecor's F-test for example.

In many cases we are not able to recognize which distribution the random sample comes from. In such cases we can use nonparametric approach. However, there are often asymptotic theorems used as well which tell about limit distribution of an estimator holding under some assumptions.

Another case is if we assume some type of the distribution and use some estimation method based on this distribution and then ask what if our assumption about the true type of distribution were wrong. Is still the used estimation method appropriate? This could be a problem of least square estimation in regression model, for example, where normal distribution of error terms is assumed and all properties of the least square estimator are examined under this assumption. Then it is convenient to ask whether the used estimator would still be optimal under small deviation from the true distribution, hence whether the estimator is "robust". From this point of view we complete the introductory definition of the model in the following way.

Let $\mathbb{X} = (X_1, \ldots, X_n)$ be a random sample from distribution P defined on σ-algebra B. Suppose that P is an element of the family of distributions $\mathcal{P} = \{P_\theta, \theta \in \Theta \subseteq \mathbb{R}^p\}$. Then by θ we will understand a functional $\theta = T(P) : \mathcal{P} \to \mathbb{R}$. The estimator of θ can then be understand as $T(P_n)$ where P_n is an empirical probability distribution of a random sample $\mathbb{X}$, i.e.

$$P_n(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{[X_i \in A]}, \quad A \in B.$$

Robustness of estimation is generally understood as an insensitivity to deviations from some assumption of the model. Generally insensitivity of $T(P_n)$ properties to small deviations from P are concerned. We will present the definition of qualitatively robust estimator as is described by Jurečková [2001]. But to do this we need some measure of a distance to be defined on $\mathcal{P}$.

Let $\mathcal{S}$ be a complete and separable metric space with a metric d and $\mathcal{B}$ be a σ-algebra of all Borel subsets of $\mathcal{S}$. Let $\mathcal{P}$ be a collection of all probability measures on $(\mathcal{S}, \mathcal{B})$. Then $\mathcal{P}$ is a convex set and we can define a distance measure for all elements $P, Q \in \mathcal{P}$. The following definition describes one type of distance measure, the Prochorov distance.

Definition 1. By *Prochorov distance* of two elements $P, Q \in \mathcal{P}$ we understand

$$d_P(P, Q) = \inf\{\epsilon > 0 : P(A) \leq Q(A^\epsilon) + \epsilon \; \forall A \in \mathcal{B}, A \neq \emptyset\},$$

where $A^\epsilon = \{x \in \mathcal{S} : \inf_{y \in A} d(x, y) \leq \epsilon\}$ is a closed ϵ-neighborhood of nonempty set A.

Let $(X_1, \dots, X_n)$ be a vector of independent identically distributed random variables with values in product measurable space $(\prod_{i=1}^n \mathcal{S}, \otimes_{i=1}^n \mathcal{B})$ and let $T_n = T_n(X_1, \dots, X_n)$ be a sequence of functionals $T_n : (\prod_{i=1}^n \mathcal{S}, \otimes_{i=1}^n \mathcal{B}) \to (\mathcal{T}_n, \mathcal{A}_n)$. Let $\mathcal{P}$ be a collection of all probability distributions on $\mathcal{B}$ with Prochorov distance.

Definition 2. The sequence of the statistics $\{T_n\}_{n\in\mathbb{N}}$ is called *robust* for probability distribution P if it holds that

$$\forall \epsilon > 0 \quad \exists \delta > 0 \text{ and } n_0 \in \mathbb{N} : \forall Q \in \mathcal{P}, \forall n \geq n_0 \text{ it holds}$$
$$d_P(P, Q) < \delta \Rightarrow d_P(\mathcal{L}_P(T_n), \mathcal{L}_Q(T_n)) < \epsilon,$$

where $\mathcal{L}_P(T_n)$ is a distribution of T_n under P, $\mathcal{L}_Q(T_n)$ is a distribution of T_n under Q and d_P is Prochorov distance.

The Definition 2 says that the estimator is robust if for any other probability measure Q that is closed in some meaning to the underlying probability measure P, the distribution of the estimator is not changed radically. The Prochorov distance itself is not essential in this definition since it can be obviously replaced by any other distance measure defined on $\mathcal{P}$.

According to Definition 2 we can say which estimator is robust and which is not, but we cannot measure the robustness. That is why it is called qualitative definition of robustness.

If we would like to measure robustness and compare estimators with each other as well, we need some quantitative characteristics of robustness. One of such measures is a tail behaviour of an estimator.

2.2 Tail behaviour of statistic estimators

Tail measure is a characteristic of robustness which is generally used when evaluating estimators of location or regression parameters. In this thesis we will concentrate on location parameters only.

Suppose that $(X_1, \dots X_n)$ is a random sample from a absolutely continuous distribution function $F(x - \theta)$ where $\theta \in \mathbb{R}$ is a location parameter we would like to estimate. In such case it is obvious that it suffices to focus only on estimators which are equivariant to location. Such estimators are called translation equivariant in literature and are described in the next definition.

Definition 3. We say that the estimator $T_n(X_1, \ldots, X_n)$ is *translation equivariant* if $\forall c \in \mathbb{R}$ and $\forall (X_1, \ldots, X_n)$ it holds that

$$T_n(X_1 + c, \ldots, X_n + c) = T_n(X_1, \ldots, X_n) + c.$$

Let us now focus on the tail behaviour of such estimators.

Definition 4. Probability $P_\theta(T_n - \theta > a)$ resp. $P_\theta(T_n - \theta < -a)$ for large $a > 0$ is called *right* resp. *left distribution tail* of T_n. In the case of a symmetric distribution by the tails of the distribution we understand $P_\theta(|T_n - \theta| > a)$ for large $a > 0$.

We could assume that good estimator should have these probabilities as low as possible, hence we should require convergence of these tails to zero. Convergence of tails can be expressed either for fixed a and $n \to \infty$, i.e.

$$\lim_{n \to \infty} P_\theta(|T_n - \theta| > a) = 0 \quad \forall a > 0,$$

which is the case of consistent estimators, or it can be expressed for fixed n and $a \to \infty$, i.e.

$$\lim_{a \to \infty} P_\theta(|T_n - \theta| > a) = 0 \quad \forall n \in \mathbb{N}.$$

Let us focus on the second type of the convergence. We could say that the faster this limit tends to zero the better is the tail behaviour of the estimator. Hence we would like to compare estimators according to the speed of this limit. Jurečková [1981] presented the following characteristic of the tail behaviour of estimators for the symmetric probability distributions.

$$B(T_n, a) = \frac{-\log P_\theta(|T_n - \theta| > a)}{-\log(1 - F(a))}, \quad a > 0.$$

Since we suppose the estimator to be translation equivariant we can also write

$$B(T_n, a) = \frac{-\log P_0(|T_n| > a)}{-\log(1 - F(a))}, \quad a > 0.$$

Remark 2.2.1. It is easy to see that

$$P_0(|T_n| > a) = (1 - F(a))^{B(T_n,a)}$$

for all $a > 0$. This equation shows that for every a the value of $B(T_n, a)$ expresses the power of a right tail of distribution F which equals the value of the distribution tails of the estimator. Hence the higher the value of $B(T_n, a)$ corresponding to some estimator T_n is, the smaller the distribution tails of T_n are for some fixed $a > 0$.

Let us fix n and suppose $a \to \infty$, then the limit of $B(T_n, a)$ tells us how much faster the convergence of log of the distribution tails of T_n is to zero than the convergence of $\log(1 - F(a))$, which is obviously the log of distribution tail of a single observation. Hence we can certainly claim the estimator with higher values (or limit) of $B(T_n, a)$ to be better. Jurečková [1981] shows that the limit of $B(T_n, a)$ for $a \to \infty$ is bounded, especially the limit cannot be larger than n and smaller then 1.

Theorem 2.2.1. Let $(X_1, \ldots, X_n)$ be a random sample from a population with the distribution function $F(x - \theta), 0 < F(x) < 1, F(-x) = 1 - F(x), x, \theta \in \mathbb{R}$. Let T_n be a translation equivariant estimator of θ such that for every $n \in \mathbb{N}_0$ it holds that

$$\min_{1\le i\le n} X_i > 0 \Rightarrow T_n(X_1, \ldots, X_n) > 0,$$
$$\max_{1\le i\le n} X_i < 0 \Rightarrow T_n(X_1, \ldots, X_n) < 0.$$

Then for every fixed $n \in \mathbb{N}$ it holds

$$1 \le \liminf_{a\to\infty} B(T_n, a) \le \limsup_{a\to\infty} B(T_n, a) \le n. \tag{2.2.1}$$

Proof. Since T_n is translation equivariant and since from the assumption of the Theorem follows for the events that $[\min_{1\le i\le n} X_i > 0] \subseteq [T_n(X_1, \ldots, X_n) > 0]$ and $[\max_{1\le i\le n} X_i < 0] \subseteq [T_n(X_1, \ldots, X_n) < 0]$, we have

$$\begin{aligned}
P_0(|T_n| > a) &= P_0(T_n > a) + P_0(T_n < -a) \\
&= P_0(T_n(X_1 - a, \ldots, X_n - a) > 0) + P_0(T_n(X_1 + a, \ldots, X_n + a) < 0) \\
&\ge P_0(\min_{1\le i\le n} X_i > a) + P_0(\max_{1\le i\le n} X_i < -a) = 2P_0(\max_{1\le i\le n} X_i < -a) \\
&= 2F^n(-a) = 2(1 - F(a))^n,
\end{aligned}$$

hence we have

$$-\log P_0(|T_n| > a) \leq -\log 2 - n\log(1 - F(a)) \leq -n\log(1 - F(a))$$

and from

$$\frac{-\log P_0(|T_n| > a)}{-\log(1 - F(a))} \leq n$$

it follows that

$$\lim_{a\to\infty} \sup B(T_n, a) \leq n.$$

Similarly, since it can be shown that from the assumption of the Theorem it follows that $[T_n(X_1, \ldots, X_n) > 0] \subseteq [\min_{1\leq i\leq n} X_i \leq 0]$ and $[T_n(X_1, \ldots, X_n) < 0] \subseteq [\max_{1\leq i\leq n} X_i \geq 0]$, we have

$$\begin{aligned}
P_0(|T_n| > a) &= P_0(T_n > a) + P_0(T_n < -a) \\
&= P_0(T_n(X_1 - a, \ldots, X_n - a) > 0) + P_0(T_n(X_1 + a, \ldots, X_n + a) < 0) \\
&\leq P_0(\min_{1\leq i\leq n} X_i \leq -a) + P_0(\max_{1\leq i\leq n} X_i \geq a) = 2P_0(\max_{1\leq i\leq n} X_i \geq a) \\
&= 2(1 - F^n(a)),
\end{aligned}$$

hence

$$-\log P_0(|T_n| > a) \geq -\log 2 - \log(1 - F^n(a))$$

and

$$\lim_{a\to\infty} \inf B(T_n, a) \geq 1.$$

□

Theorem 2.2.1 tells that the fastest distribution tails of T_n are n-times faster than log of $1 - F(a)$ hence log of the distribution tails of a single observation. On the other hand the slowest tails are as fast as the distribution tails of a single observation.

There appears a question whether some estimator can have large values of $B(T_n, a)$ and if it can hold for a large family of distribution functions at the same time. It is obvious that the value of $B(T_n, a)$ depends not only on given estimator T_n but also on the tail behaviour of the distribution F. The next definition distinguishes between two types of behaviour of distribution tails.

Definition 5. Let the distribution function F satisfy $F(-x) = 1 - F(x)$, $x \in \mathbb{R}$. We say that the distribution function F has an *exponential tails* (light tails, type I. tails) if it holds that

$$\lim_{a \to \infty} \frac{-\log(1 - F(a))}{ba^r} = 1, \quad b > 0, r \geq 1.$$

We say that the distribution function F has an *algebraic tails* (heavy tails, type II. tails) if it holds that

$$\lim_{a \to \infty} \frac{-\log(1 - F(a))}{m \log a} = 1, \quad m > 0.$$

Remark 2.2.2. Examples of distributions with exponential tails are normal ($r = 2$), logistic or Laplace ($r = 1$). Distributions with heavy tails are for example Cauchy ($m = 1$) or t-distribution with m degrees of freedom ($m > 1$).

Definition 5 distinguishes distribution functions according to the speed of the convergence of corresponding distribution tails. Whereas exponential tails tends to zero for $a \to \infty$ at the same speed as an exponential function, heavy tails tends to zero slower, at the same speed as algebraic polynomial.

If we consider only this division of tail behaviour of distribution functions, then our demand on $B(T_n, a)$ values is restricted to the question whether it is possible to have estimators with high values of $B(T_n, a)$ for both exponential and heavy tailed distributions. It will be shown that among monotone and translation equivariant estimators we cannot find such estimator. Monotone and translation equivariant estimators which have stable values of $B(T_n, a)$ for both light or heavy tailed distributions have moderately converging tails in sense that they do not reach neither upper bound nor lower bound in (2.2.1). On the other hand estimators with fast converging tails can reach upper bound in (2.2.1) only for light tailed distributions, for heavy tailed distributions on the other hand their tails converge very slow.

The tail behaviour of estimators as a measure of robustness is in very closed relationship to another measure used for estimators evaluation, the breakdown point.

2.3 Breakdown point

The concept of breakdown points was researched by several authors in last few decades and a few alternative definitions were introduced. For the first time it was introduced by Donoho and Huber in 1983 and was refered to finite sample. Since this finite sample breakdown point generally depends on the size of the sample, asymptotic breakdown point was introduced as a limit of a breakdown point as the sample size tends to infinity. The finite sample breakdown point is defined as follows.

Let $\mathbf{x}^0 = (x_1, \ldots, x_n)$ be an initial random sample and $T_n(\mathbf{x}^0)$ is a value of the estimator in $\mathbf{x}^0$. Denote $\mathbf{x}^m$ a contaminated random sample $\mathbf{x}^0$ which is created by replacing m arbitrary elements of $\mathbf{x}^0$ with any other arbitrarily chosen values. Let $T_n(\mathbf{x}^m)$ be the value of the estimator in $\mathbf{x}^m$.

Definition 6. The number

$$\epsilon_n^*(T_n, \mathbf{x}^0) = \frac{m^*(\mathbf{x}^0)}{n},$$

where $m^*(\mathbf{x}^0)$ is the lowest number $m \in \mathbb{N}_0$, such that

$$\sup_{\mathbf{x}^m} \| T_n(\mathbf{x}^m) - T_n(\mathbf{x}^0) \| = \infty,$$

is called the finite sample *breakdown point* of the estimator T_n for a random sample $\mathbf{x}^0$.

Remark 2.3.1. As will be shown under some assumptions $m^*(\mathbf{x}^0)$ does not depend on the initial sample $\mathbf{x}^0$. Then we can set a limit

$$\epsilon^* = \lim_{n \to \infty} \epsilon_n^*,$$

which is sometimes called breakdown point as well. In this thesis such limit will be called asymptotic breakdown point.

The value of $m^*(\mathbf{x}^0)$ in definition 6 represent the smallest fraction of a sample $\mathbf{x}^0$ which after replacing with arbitrary values can cause the estimator to take on values arbitrarily far from the initial value of the estimator $T_n(\mathbf{x}^0)$. Hence the value of a breakdown point represents the part of observations which replacing by another unfavorable values can cause a failure of the estimator. This measure of robustness is helpful in telling how much the estimator is useful when probability distribution we make a random sample from is mixed with another distribution (for example heavy tailed, which can cause outliers to be presented in our sample). In case of such mixture we say that the probability distribution is contaminated by another distribution how the next definition shows.

Definition 7. Let $P, Q \in \mathcal{P}$ and $t \in [0, 1]$. The probability distribution of

$$P_t(Q) = (1-t)P + tQ$$

is called the *contamination* of P by distribution Q in proportion t.

Remark 2.3.2. $P_0(Q)$ means no contamination of P by distribution Q whereas $P_1(Q)$ means full contamination.

For contaminated distributions following important Lemma holds as was shown by Jurečková [1981].

Lemma 2.3.1. Let $F(x) = (1-t)G(x) + tH(x)$, where G and H are absolutely continuous distribution functions with symmetric positive densities g and h and $0 < t < 1$. If

$$\lim_{x\to\infty} \frac{1-G(x)}{1-H(x)} = 0$$

and

$$\lim_{x\to\infty} \frac{g(x)}{h(x)} = 0$$

then

$$\lim_{x\to\infty} \frac{\log(1-F(x))}{\log(1-H(x))} = 1.$$

Proof.

$$\lim_{x\to\infty} \frac{\log(1-F(x))}{\log(1-H(x))} = \lim_{x\to\infty} \frac{(1-H(x))f(x)}{(1-F(x))h(x)}$$

$$= \lim_{x\to\infty} \frac{(1-H(x))((1-t)g(x)+th(x))}{(1-(1-t)G(x)-tH(x))h(x)} = \lim_{x\to\infty} \frac{(1-t)\frac{g(x)}{h(x)}+t}{(1-t)\frac{1-G(x)}{1-H(x)}+t} = 1.$$

□

Lemma 2.3.1 tells that if the distribution F is a mixture of two distributions G and H and tails of G converges to zero faster than tails of H, then tails of F converges to zero in the same speed as H, hence as the tails of distribution with slower tails. This means that whenever we make a random sample from a distribution though with exponential tails which is contaminated by some heavy tailed distribution, the resulting distribution is always heavy tailed. This is an important fact since it can have significant effect on used estimator. Suppose simple case where a random sample is made from the normal distribution. When the distribution is not contaminated, the sample mean is the best estimator of location parameter in sense we have mentioned before that its distribution tails converge to zero the fastest they can for fixed n and $a \to \infty$. On the other hand when the initial distribution is contaminated by some heavy tailed distribution, then the resulting distribution is heavy tailed as well and the sample mean becomes the worst estimator in sense that its distribution tails converge to zero the slowest they can, i.e. as distribution tails of a single observation.

The role of breakdown point then is to show how at least the original distribution has to be contaminated by some unfavorable distribution to cause the estimator to fail. It is, expectedly, in a close relationship with the tail behaviour of the estimator distribution. This issue was closely dealt by He, Jurečková, Koekner and Portnoy [1990]. Before we show this relationship in Theorem 2.3.1, we introduce lemma which shows that for some types of estimators the value of $m^*(\mathbf{x}^0)$ does not depend on the initial observations $\mathbf{x}^0$.

Lemma 2.3.2. Suppose $T_n(X_1, \ldots, X_n)$ is a translation equivariant estimator of θ such that T_n is nondecreasing in each argument X_i. Then T_n has a universal breakdown point, i.e. $m^*(\mathbf{x}^0) = m^*$ for all initial vectors of observations $\mathbf{x}^0$.

Proof. We will show that under assumptions of Lemma it holds that $m^*(\mathbf{x}) = m^*(\mathbf{0}_n)$, where $\mathbf{0}_n$ denote a vector of n zeros, for any initial sample $\mathbf{x}$.

Suppose $m^*(\mathbf{x}) = m$ and set $c = \max_{i=1,\ldots,n} |x_i|$. From definition of m^* it holds that for every $B > 0$ there exists a contaminated sample $z = (y, x)$, where y is a vector of m chosen values and x is a vector of $n - m$ elements from $\mathbf{x}$, such that $|T_n(z)| > B + c$. If $T_n(z) > 0$ then $T_n(z - c) > B$ which is $T_n(y - c, x - c) > B$.

From monotony of T_n we get $T_n(y-c, \mathbf{0}_{n-m}) \geq T_n(y-c, x-c) > B$. If $T_n(z) > 0$ then $T_n(z+c) < -B$ and from monotony it holds $T_n(y+c, \mathbf{0}_{n-m}) \leq T_n(y+c, x+c) < -B$. Hence for every $B > 0$ there exists a random sample z of n zeros contaminated by m arbitrary values such that $|T_n(z)| > B$.

Furthermore, from definition of m^* it follows that there exists $B > 0$ such that for all contaminated samples $z = (y, x)$, where y is a vector of $m-1$ chosen values and x is a vector of $n-m+1$ elements from $\mathbf{x}$, such that $|T_n(z)| < B - c$. By similar process we get that it follows from here that for every sample z of n zeros contaminated by $m-1$ values it holds that $|T_n(z)| < B$.

From here it follows that $m^*(\mathbf{0}_n) = m$. $\square$

Theorem 2.3.1. Suppose $T_n(X_1, \ldots, X_n)$ is a translation equivariant estimator of θ such that T_n is nondecreasing in each argument X_i. Then for any symmetric, absolutely continuous F with density $f(z) = f(-z) > 0, z \in \mathbb{R}$, and such that

$$\lim_{z\to\infty} \frac{\log(1-F(z+c))}{\log(1-F(z))} = 1, \quad \text{for any fixed } c > 0,$$

if holds that

$$m^* \leq \lim_{a\to\infty} \inf B(a, T_n) \leq \lim_{a\to\infty} \sup B(a, T_n) \leq n - m^* + 1, \tag{2.3.1}$$

where m^*/n is the universal breakdown point of T_n.

For proof of this Theorem we use following Lemma.

Lemma 2.3.3. Suppose $T_n(X_1, \ldots, X_n)$ is a translation equivariant estimator of θ such that T_n is nondecreasing in each argument X_i. Denote $x_{(i)}$ the i-th order statistic of the sample $(x_1, \ldots, x_n)$. Then there exists constant A such that

$$x_{(m^*)} - A \leq T_n \leq x_{(n-m^*+1)} + A.$$

Proof of Lemma 2.3.3. Denote R_i the rank of x_i among $x_1, \ldots, x_n$. Then it holds that

$$\begin{aligned}
T_n(x_1, \ldots, x_n) &= T_n(x_1 - x_{(m^*)}, \ldots, x_n - x_{(m^*)}) + x_{(m^*)} \\
&\geq T_n(x_{R_1} - x_{(m^*)}, \ldots, x_{R_n} - x_{(m^*)}) + x_{(m^*)} \\
&\geq T_n((x_{R_1} - x_{(m^*)})\mathbb{I}_{[R_1 \leq m^*]}, \ldots, (x_{R_n} - x_{(m^*)})\mathbb{I}_{[R_n \leq m^*]}) + x_{(m^*)}
\end{aligned}$$

$T_n((x_{R_1} - x_{(m^*)})\mathbb{I}_{[R_1 \le m^*]}, \ldots, (x_{R_n} - x_{(m^*)})\mathbb{I}_{[R_n \le m^*]})$ is an estimator based on a sample with $m^* - 1$ nonzero values and hence is bounded. From here it follows

$$T_n(x_1, \ldots, x_n) \ge x_{(m^*)} - A.$$

Similarly

$$\begin{aligned}
T_n(x_1, \ldots, x_n) &= T_n(x_1 - x_{(n-m^*+1)}, \ldots, x_n - x_{(n-m^*+1)}) + x_{(n-m^*+1)} \\
&\le T_n(x_{n-R_1+1} - x_{(n-m^*+1)}, \ldots, x_{n-R_n+1} - x_{(n-m^*+1)}) + x_{(n-m^*+1)} \\
&\le T_n((x_{n-R_1+1} - x_{(n-m^*+1)})\mathbb{I}_{[R_n \le m^*]}, \ldots, (x_{n-R_n+1} - x_{(n-m^*+1)})\mathbb{I}_{[R_1 \le m^*]}) \\
&\quad + x_{(n-m^*+1)}
\end{aligned}$$

and T_n in the last expression is based on sample with $m^* - 1$ nonzero values and hence is bounded. From here it follows that

$$T_n(x_1, \ldots, x_n) \ge x_{(n-m^*+1)} + A.$$

$\square$

Proof of Theorem 3.1.1. From Lemma 2.3.3 it follows that

$$\begin{aligned}
P_0(T_n > a) &\ge P_0(X_{(m^*)} > a + A) \ge P_0(X_1 > a + A, \ldots, X_{n-m^*+1} > a + A) \\
&= (1 - F(a + A))^{n-m^*+1}
\end{aligned}$$

Hence we get

$$\frac{\log 2P_0(T_n > a)}{-\log(1 - F(a))} \le \frac{-\log 2 + (n - m^* + 1)\log(1 - F(a + A))}{\log(1 - F(a))}$$

From here and assumption of the Theorem we get

$$\limsup_{a \to \infty} B(T_n, a) \le n - m^* + 1.$$

Similarly,

$$\begin{aligned}
P_0(T_n > a) &\le P_0(X_{(n-m^*+1)} > a + A) \\
&= n\binom{n-1}{m^*-1} \int_{a+A}^{\infty} F(x)^{n-m^*}(1 - F(x))^{m^*-1} f(x)dx \\
&\le n\binom{n-1}{m^*-1} \int_{F(a+A)}^{\infty} (1 - t)^{m^*-1} dt \\
&= \binom{n}{m}(1 - F(a + A))^{m}.
\end{aligned}$$

Hence

$$\frac{-\log 2P_0(T_n > a)}{-\log(1-F(a))} \geq \frac{\log 2\binom{n}{m} + (m^*)\log(1-F(a+A))}{\log(1-F(a))}$$

Hence it follows that

$$\lim_{a\to\infty} \inf B(T_n, a) \geq m^*.$$

□

Theorem 2.3.1 shows fundamental relationship between the tail behaviour of estimators and their breakdown points for monotone and translation equivariant estimators. It can be seen that the higher the breakdown point of the estimator is, i.e. the more the estimator is robust to the contamination by heavy tailed distribution, the lower the upper bound in (2.3.1) is. Hence the more the estimator is robust in the sense of the high breakdown point, the slower its distribution tails are able to converge to zero for $a \to \infty$. On the other hand, estimators with higher breakdown have higher lower bound in (2.3.1) hence high breakdown point estimators cannot have slowest tails for any distribution F, light or heavy tailed either. The fact that the high breakdown point estimators are maximizing the lower bound in (2.3.1) is called the minimax property.

However, this property of breakdown point is well established only for monotone and translation equivariant estimators. Whether the inequalities (2.3.1) can be generalized also for nonmonotone estimators was dealt by Kušnier and Mizera [2001] concerning the scale equivariant estimators, i.e. estimators satisfying $T_n(cX_1, \ldots, cX_n) = cT_n(X_1, \ldots, X_n)$ for any $c \in \mathbb{R}$. The main conclusion of their research is that under more general assumptions there exist translation and scale equivariant estimators such that they show worse behaviour then the lower bound in (2.3.1).

Chapter 3

Robust characteristics of estimators

When estimating location parameters asymptotic theory has a strong position in estimating process. From Central Limit Theorem it holds that the sample mean has asymptotically normal distribution no matter what type the underlying distribution is. The only assumption is finite variance. Similarly, it was shown that many robust estimators have asymptotic normal distribution. Since normal distribution is light tailed, the problem of estimators' sensitivity to heavy tailed distribution does not seem to play an important role. However, in reality the sample size is never infinite and hence the true distribution never reaches the normal one perfectly (except the case when the finite sample distribution of the estimator is normal). Moreover, the critical bounds of asymptotic test depend critically on the tails of the estimator distribution and if for any finite sample the estimator distribution is heavy tailed our test could be distorted. In the next chapter it will be shown that the finite sample distribution of the translation equivariant estimator is always heavy tailed whenever the random sample comes from the heavy tailed distribution.

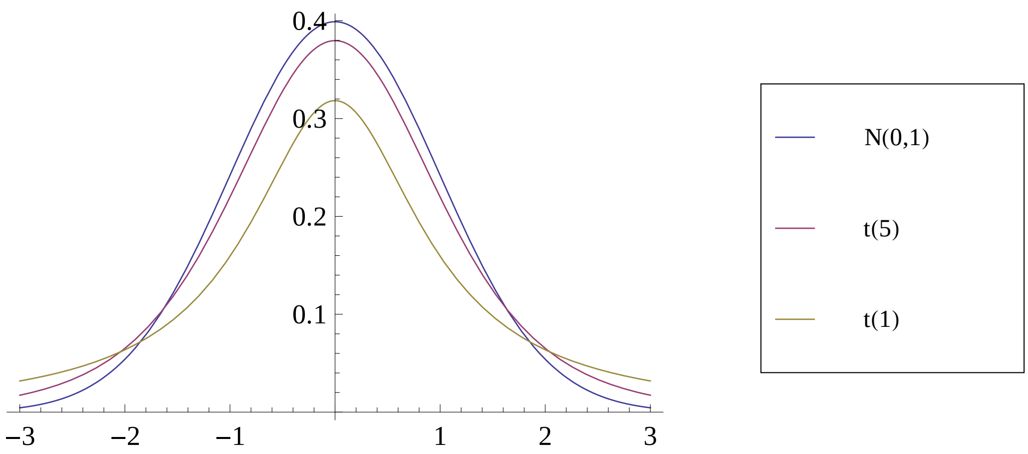
In this chapter we investigate some estimators according to the properties mentioned in the previous section, i.e. the tail behaviour and the breakdown point. We will consider only translation equivariant estimators of location parameters, hence we restrict to symmetric densities only.

Jurečková an Picek [2011] illustrated the tail behavior of some estimators under heavy tailed Cauchy distribution in comparison to the light tailed normal distribution. In this chapter we will follow and show the behaviour of estimators under Students t-distribution.

Student's t-distribution with m degrees of freedom have a density in the form

$$f(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\sqrt{\pi m}}\left(1+\frac{x^2}{m}\right)^{-\frac{m+1}{2}}.$$

This distribution is heavy tailed for any finite m. The special case of t-distribution for $m = 1$ is a Cauchy distribution. However, for m going to infinity density of the t-distribution approaches density of the normal distribution hence the one with light tails. Graph 1 compares densities of the standard normal distribution with the Student t-distribution with 1 and 5 degrees of freedom.



Graph 1: Densities of N(0,1) distribution and Student t-distribution with 1 and 5 degrees of freedom.

In this chapter we illustrate how the tails of estimators behave when sample comes from the t-distribution with 1 to 5 degrees of freedom. As will be shown in the third chapter, distribution of discussed estimators is heavy tailed for any finite n. Hence furthermore, we make numerical illustrations describing approximations of the distribution of estimators by the normal or t-distribution.

3.1 Sample mean

Tail behaviour of sample mean was firstly shown by Jurečková [1979], where besides it was compared to the tail behaviour of L-estimators there were shown limit of characteristic $B(T_n, a)$ when the initial sample comes from heavy tailed distribution. Jurečková [1981] followed and showed the tail behaviour of sample mean when the initial sample comes from light tailed distribution.

Consider $X_1, \dots X_n$ to be a random sample from the population with the distribution function $F(x-\theta)$ and density $f(x-\theta)$ such that $f(x) = f(-x)$ where $x \in \mathbb{R}$ and $\theta \in \mathbb{R}$ is the location parameter we would like to estimate.

Let $T_n(X_1, \dots, X_n) = \bar{X}_n$ be the estimator of θ where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ is the sample mean. Sample mean is very often used as a location estimator because of its pleasant properties as far as the normal distribution is concerned. We know that if the distribution function F is normal then the distribution of $\bar{X}_n$ is normal as well. On the other hand even in the case we do not know the true distribution F, according to the Central Limit Theorem $\bar{X}_n$ has asymptotically normal distribution if an initial random sample has a finite variance. These properties are very often used for testing hypothesis about θ. Let us focus on robust characteristics of sample mean.

Theorem 3.1.1. Let $X_1, \dots X_n$ be the random sample from distribution function F with symmetric density f. Then for the sample mean $\bar{X}_n$ it holds

i) If the distribution function F has light tails, then

$$\lim_{a\to\infty} B(a, \bar{X}_n) = n.$$

ii) If the distribution function F has heavy tails, then

$$\lim_{a\to\infty} B(a, \bar{X}_n) = 1.$$

For proof of Theorem 3.1.1 we will use the following Lemma.

Lemma 3.1.1. Let $(X_1, \dots, X_n)$ be a random sample from distribution with distribution function $F(x-\theta)$ and density $f(x-\theta)$ such that $f(-x) = f(x) > 0$. Let T_n be any translation equivariant estimator of θ and d_n any positive number such that

$$E_0[\exp\{d_n b|T_n|^r\}] < \infty,$$

where $b > 0$ and $r \geq 1$. Then it holds that

$$\lim_{a\to\infty} \inf \left[\frac{-\log P_0(|T_n| > a)}{ba^r}\right] \geq d_n$$

Proof of Lemma 3.1.1. Since $\exp\{d_n b a^r\}$ is nondecreasing function of a we can write

$$P_0(|T_n| > a) = P_0\left[\exp\{d_n b |T_n|^r\} > \exp\{d_n b a^r\}\right].$$

From Markov inequality moreover

$$P_0\left[\exp\{d_n b |T_n|^r\} > \exp\{d_n b a^r\}\right] \leq \frac{E_0[\exp\{d_n b |T_n|^r\}]}{\exp\{d_n b a^r\}}$$

hence

$$\begin{aligned}\liminf_{a\to\infty} & \left[\frac{-\log P_0(|T_n| > a)}{b a^r}\right] \\ & \geq \liminf_{a\to\infty}\left[\frac{-\log E_0[\exp\{d_n b|T_n|^r\}]}{b a^r} + d_n\right] = d_n.\end{aligned}$$

□

Proof of Theorem 3.1.1. Since F is light tailed, it holds that

$$\lim_{a\to\infty}\frac{-\log 2(1-F(a))}{b a^r} = 1, \quad b > 0, r \geq 1$$

and from here it follows that for any $\epsilon \in (0,1)$ we can find $K_\epsilon > 0$ such that for all $a \geq K_\epsilon$ it holds that

$$\begin{gathered} b a^r - \frac{\epsilon}{2} b a^r \leq -\log 2(1-F(a)) \leq b a^r + \frac{\epsilon}{2} b a^r, \\ \exp\left(b a^r\left(-1+\frac{\epsilon}{2}\right)\right) \geq 2(1-F(a)) \leq \exp\left(b a^r\left(-1-\frac{\epsilon}{2}\right)\right), \\ 1-\exp\left(b a^r\left(-1-\frac{\epsilon}{2}\right)\right) \leq 2F(a)-1 \leq 1-\exp\left(b a^r\left(-1+\frac{\epsilon}{2}\right)\right). \end{gathered}$$

Let L_ϵ be the largest number from $[0, K_\epsilon]$ such that

$$1-\exp\left(b L_\epsilon^r\left(-1-\frac{\epsilon}{2}\right)\right) = 2F(L_\epsilon)-1.$$

Such number always exists since the equality holds always for 0. Consider the distribution function

$$G_\epsilon(x) = \begin{cases} 0 & \text{for } x \leq 0, \\ 2F(x)-1 & \text{for } 0 < x \leq L_\epsilon, \\ 1-\exp\left(b x^r\left(-1-\frac{\epsilon}{2}\right)\right) & \text{for } L_\epsilon < x. \end{cases}$$

Put $d_n = (1-\epsilon)n$. Then from Hölder's inequality it holds that

$$E_0\left[\exp{(1-\epsilon)nb|\bar{X}_n|^r}\right] = E_0\left[\exp{(1-\epsilon)n^{1-r}b|\sum_{i=1}^{n}X_i|^r}\right]$$
$$\leq E_0\left[\exp{(1-\epsilon)b\sum_{i=1}^{n}|X_i|^r}\right] = \left(E_0\left[\exp{(1-\epsilon)b\,|X_1|^r}\right]\right)^n$$

Moreover

$$E_0\left[\exp{(1-\epsilon)b\,|X_1|^r}\right] = 2\int_0^{\infty}\exp{(1-\epsilon)bx^r}dF_0(x) \leq \int_0^{L_\epsilon}\exp{(1-\epsilon)bx^r}dG_\epsilon(x)$$
$$= 2\int_0^{L_\epsilon}\exp{(1-\epsilon)bx^r}dF(x)$$
$$+ br\left(1-\frac{\epsilon}{2}\right)\int_{L_\epsilon}^{\infty}\exp{-\frac{\epsilon}{2}bx^r x^{r-1}}dx < \infty,$$

where the inequality follows from the fact that $2F(x) \geq G_\epsilon(x)+1$ for $x > 0$. Since assumption of Lemma 3.1.1 is satisfied it holds that

$$\lim_{n\to\infty}\inf\left[\frac{-\log P_0(|X_n| > a)}{ba^r}\right] \geq (1-\epsilon)n.$$

Since this holds for any $\epsilon \in (0,1)$, obviously the first equality of Theorem holds.

The second equality of Theorem can be shown by the fact that

$$P_0(|\bar{X}_n| > a) = P_0(\bar{X}_n > a) + P_0(\bar{X}_n < -a)$$
$$\geq P_0(X_1 > -a, \ldots, X_{n-1} > -a, X_n > (2n-1)a)$$
$$+ P_0(X_1 < a, \ldots, X_{n-1} < a, X_n < -(2n-1)a)$$
$$= 2(F(a))^{n-1}[1 - F((2n-1)a)]$$

Hence

$$-\log P_0(|\bar{X}_n| > a) \leq -\log 2 - (n-1)\log F(a) - \log[1 - F((2n-1)a)]$$

and furthermore

$$\lim_{n\to\infty}\sup B(a,\bar{X}_n) \leq \lim_{n\to\infty}\sup\left[\frac{-\log 2}{-\log(1-F(a))} - \frac{(n-1)\log F(a)}{-\log(1-F(a))}\right.$$
$$\left. + \frac{-\log[1-F((2n-1)a)]}{m\log[(2n-1)a]}\left(\frac{m\log(2n-1)}{-\log(1-F(a))} + \frac{m\log(a)}{-\log(1-F(a))}\right)\right] = 1.$$

$\square$

Theorem 3.1.1 tells that the sample mean has the fastest tails in sense of (2.2.1) when the random sample has light tailed distribution. In such case the tails of the distribution of $\bar{X}_n$ converge to zero for $a \to \infty$ n times faster than log of the tails of the distribution of a single observation. On the other hand when the initial distribution is heavy tailed then the sample mean is very poor estimator of θ since the tails of its distribution behave as of the single observation. From this follows that when the initial distribution is heavy tailed the distribution of sample mean is heavy tailed as well.

As far as the breakdown point is concerned it is obvious that it is enough to replace just one observation with infinite value and sample mean reaches infinity as well. Hence the breakdown point of sample mean is $1/n$. Hence one outlier in the observations can cause the sample mean to fail. The asymptotic breakdown point is then 0.

However, even though the initial distribution is heavy tailed the central limit theorem holds of course but for the number of observation going to infinity. But whenever having some random sample its size is finite and for finite number of observations the distribution of sample mean behaves as a single observation and hence is still heavy tailed. This fact can cause the critical region bounds based on asymptotically normal distribution to be distorted.

We will illustrate the tail behaviour of sample mean when random sample is from the t-distribution.

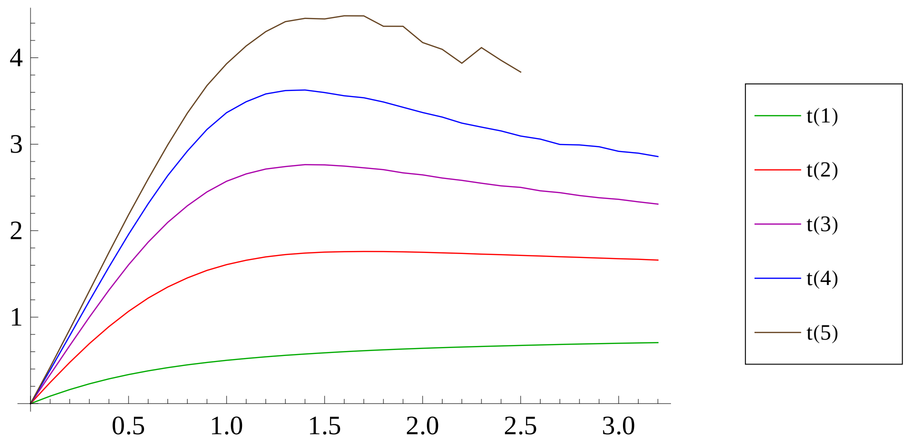
Illustration for Student t-distribution:

Let $X_1, ..., X_n$ be a random sample from population with Student t-distribution having k degrees of freedom and $\bar{X}_n$ sample mean of this random sample. We have generated 10^6 of such samples for $k = 1, \ldots, 5$ and for $n = 20$ and $n = 100$ and calculated its corresponding sample means.

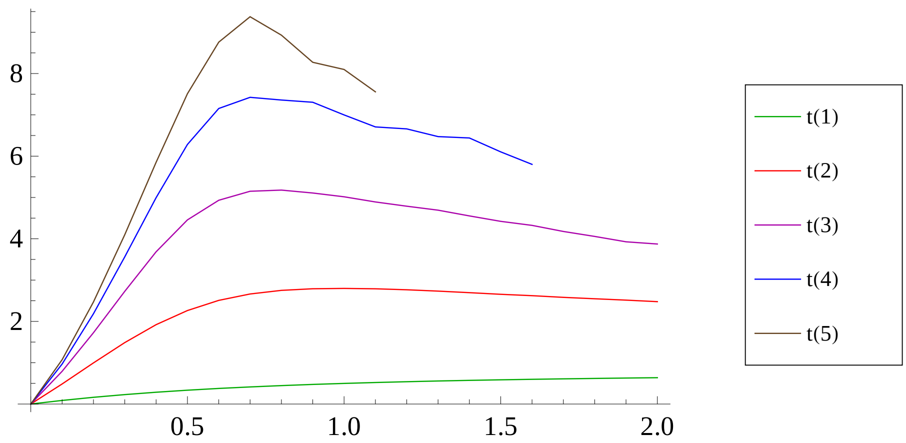
From Theorem 3.1.1 we know that

$$\lim_{a \to \infty} B(a, \bar{X}_n) = 1$$

for any finite n. However, since Student t-distribution approaches normal distribution for k tending to infinity, we expect this limit to be slower for higher k for some fixed n and moreover to be slower for higher n for some fixed $k > 2$[1]. Graph 2 and 3 shows values of $B(a, \bar{X}_n)$ for Student t-distribution with 1 to 5 degrees of freedom.



Graph 2: Values of $B(a, \bar{X}_n)$ for Student t-distribution with 1 to 5 degrees of freedom, $n = 20$.



Graph 3: Values of $B(a, \bar{X}_n)$ for Student t-distribution with 1 to 5 degrees of freedom, $n = 100$.

Although the extreme values of T_n are hard to simulate we can see the convergence of tails even for small a.

Simulations confirmed expected tail behaviour of the sample mean. For all mentioned degrees of freedom $B(a, \bar{X}_n)$ tends to 1. However, the more degrees of freedom the Student's distribution has the slower the limit is. For 1 degree

[1] We mention $k > 2$ since the finite variance in Central Limit Theorem is required.

of freedom the values of $B(a, \bar{X}_n)$ are not higher than 1 for any $a > 0$ and are very fast approaching 1 hence the tails of the sample mean are decreasing in similar speed to the tails of one observation. The more degrees of freedom the t-distribution has the higher values $B(a, \bar{X}_n)$ reaches for small a and the slower than converges to 1. From here it also follows that the more degrees of freedom the closer the distribution of sample mean is to the light tailed in the central part of the distribution. Distant tails but still remains heavy. Moreover, it can be seen that for larger sample size $B(a, \bar{X}_n)$ reaches higher values for some fixed number of degrees of freedom but higher than 1.

Since the Student t-distribution approaches normal distribution for k tending to infinity, there is a good reason to approximate the distribution of sample mean by normal one also for small sample sizes, i.e. where we do not expect the central limit theorem to be useful. However, the tails of sample mean are still heavy and hence there remains a question how much degrees of freedom we need for such approximation to be good.

Denote S_n the variance of sample mean computed from the simulation and consider standardized version of sample mean $\bar{X}_n^s$ calculated as $\frac{\bar{X}_n}{\sqrt{S_n}}$. Tables 1 and 2 compares tails of such standardized estimator with tails of standardized normal distribution and t-distribution with $n-1$ degrees of freedom. There are approximate values of $P_{t_k}(|\bar{X}_n^s| > q)$ based on simulations where q are quantiles of normal or t-distribution with $n-1$ degrees of freedom.

$t_k \diagdown q$	$u_{1-\frac{\alpha}{2},\alpha=0.05}$	$t_{1-\frac{\alpha}{2},\alpha=0.05}$	$u_{1-\frac{\alpha}{2},\alpha=0.01}$	$t_{1-\frac{\alpha}{2},\alpha=0.01}$
t_1	0.000152	0.000143	0.000113	0.000098
t_2	0.009037	0.007797	0.005030	0.004012
t_3	0.047747	0.037228	0.015719	0.009813
t_4	0.050983	0.038383	0.013293	0.007021
t_5	0.050966	0.037772	0.011929	0.005860

Table 1: Simulated values of $P_{t_k}(|\bar{X}_n^s| > q)$ for $n = 20$.[2]

[2] u_β is a β quantile of $N(0,1)$ distribution and t_β is β quantile of Student t-distribution with

$t_k \diagdown q$	$u_{1-\frac{\alpha}{2},\alpha=0.05}$	$t_{1-\frac{\alpha}{2},\alpha=0.05}$	$u_{1-\frac{\alpha}{2},\alpha=0.01}$	$t_{1-\frac{\alpha}{2},\alpha=0.01}$
t_1	0.000133	0.000121	0.000102	0.000090
t_2	0.018873	0.015965	0.009544	0.007428
t_3	0.048696	0.036734	0.013018	0.007108
t_4	0.050149	0.036873	0.010926	0.005146
t_5	0.050308	0.036893	0.010462	0.004654

Table 2: Simulated values of $P_{t_k}(|\bar{X}_n^s| > q)$ for $n = 100$.[3]

For $n = 20$ the simulated tails of sample mean are good approximated by tails of normal distribution for degrees of freedom higher or equal to 4 for $\alpha = 0.05$ and rather higher or equal to 5 for $\alpha = 0.01$. For $n = 100$ the tails of sample mean are quite well approximated by tails of normal distribution for degrees of freedom higher or equal to 3 for $\alpha = 0.05$ and rather higher or equal to 4 for $\alpha = 0.01$.

Table 3 shows simulated 0.975 quantiles of standardized sample mean.

t_k	$q_{0.975}, n = 20$	$q_{0.975}, n = 100$
t_1	0.00542	0.00476
t_2	0.95542	1.36684
t_3	1.93887	1.94634
t_4	1.97043	1.95844
t_5	1.96754	1.96475
t_6	1.96806	1.96221

Table 3: Simulated values of 0.975 quantiles of standardized sample mean for $n = 20$ and $n = 100$ and t-distribution with k degrees of freedom.

$n - 1$ degrees of freedom.

[3] u_β is a β quantile of $N(0, 1)$ distribution and t_β is β quantile of Student t-distribution with $n - 1$ degrees of freedom.

3.2 L-estimators

Tail behaviour of L-estimators was firstly described by Jurečková [1979] and followed by Jurečková [1981].

Suppose again $X_1, \dots X_n$ to be a random sample from the population with the distribution function $F(x-\theta)$ and density $f(x-\theta)$ such that $f(x) = f(-x)$ where $x \in \mathbb{R}$ and $\theta \in \mathbb{R}$ is the location parameter.

Let T_n be an estimator of parameter θ of the form

$$T_n = \sum_{i=1}^{n} c_i X^{(i)},$$

where $X^{(1)} \leq ... \leq X^{(n)}$ are the order statistics corresponding to $X_1, ..., X_n$ and $c_i \geq 0$ are coefficients such that $c_i = c_{n-i+1}$ for $i = 1, ..., n$ and it holds that $\sum_{i=1}^{n} c_i = 1$. Such estimator is called L-estimator and is obviously translation equivariant. Following theorem shows properties of its tail behavior.

Theorem 3.2.1. Let $(X_1, \dots, X_n)$ be a random sample from distribution with distribution function $F(x-\theta)$ and density $f(x-\theta)$ such that $f(-x) = f(x) > 0$. Let T_n be an L-estimator of parameter θ and assume that $c_i = c_{n-i+1} = 0$ for $i = 0, 1, ..., k$ where $0 \leq k < \frac{n}{2}$. Then it holds that

$$k + 1 \leq \lim_{a\to\infty} \inf B(T_n, a) \leq \lim_{a\to\infty} \sup B(T_n, a) \leq n - k \tag{3.2.1}$$

Proof. From the form of estimator T_n we have

$$\begin{aligned}
P_0(|T_n| > a) &= P_0(T_n > a) + P_0(T_n < -a) \leq P_0(X^{(n-k)} > a) + P_0(X^{(k+1)} < -a) \\
&= n\binom{n-1}{k} \int_a^\infty F(x)^{n-k}(1 - F(x))^k f(x)dx \\
&\qquad + n\binom{n-1}{k} \int_{-\infty}^{a} F(x)^k (1 - F(x))^{n-k-1} f(x)dx \\
&= 2n\binom{n-1}{k} \int_0^{1-F(a)} t^k(1-t)^{n-k-1}dt \leq 2n\binom{n-1}{k} \int_0^{1-F(a)} t^k dt \\
&= 2\binom{n}{k+1}(1 - F(a))^{k+1},
\end{aligned}$$

where for the first integral we used substitution $1 - F(x) = t$ and for the second integral substitution $F(x) = t$. Moreover we used the fact that $(1-t)^{n-k-1} < 1$ for $t \in (0, 1 - F(a))$. From here it follows that

$$-\log P_0(|T_n| > a) \geq -(k + 1)\log(1 - F(a))$$

and hence

$$\lim_{a\to\infty}\inf B(T_n,a)=\lim_{a\to\infty}\sup\frac{-\log P_0(|T_n|>a)}{-\log(1-F(a))}\geq k+1.$$

From the form of estimator T_n it also holds that

$$\begin{aligned}P_0(|T_n|>a)&=P_0(T_n>a)+P_0(T_n<-a)\geq P_0(X^{(k+1)}>a)+P_0(X^{(n-k)}<-a)\\&=2n\binom{n-1}{k}\int_0^{1-F(a)}t^{n-k-1}(1-t)^k dt\\&\geq 2n\binom{n-1}{k}\int_0^{1-F(a)}t^{n-k-1}(F(a))^k dt\\&=2\binom{n}{k}(F(a))^k(1-F(a))^{n-k},\end{aligned}$$

where we use the same substitutions as before and also the fact that $(1-t)^k\geq(F(a))^k$ for $t\in(0,1-F(a))$. From here it follows that

$$-\log P_0(|T_n|>a)\leq-\log\left[2\binom{n}{k}\right]-(n-k)\log(1-F(a))$$

and hence it holds that

$$\lim_{a\to\infty}\sup B(T_n,a)=\lim_{a\to\infty}\sup\frac{-\log P_0(|T_n|>a)}{-\log(1-F(a))}\leq n-k.$$

□

Obviously the constant k in Theorem 3.2.1 represents the part of observations which are cut off before computing estimator (exactly $2k$ observations are cut off). Cutting off extreme observations is called trimming and from Theorem 3.2.1 it then follows that the higher number of extreme observations are trimmed off, the more robust the estimator is in sense that the closer are the bounds in (3.2.1) to each other. On the other hand the lower k is the more the tail behaviour is similar to the sample mean.

As far as the breakdown point of L-estimators is concerned it is obvious that at least one element with weight unequal to zero of a random sample has to be replaced with infinity value to cause the L-estimator to fail. If we will consider k the number of trimmed off values as mentioned in Theorem 3.2.1 then the minimum elements needed to be replaced by infinity values to cause the estimator to fail is $k+1$. Hence it holds that the breakdown point of L-estimators is

$$\epsilon^*(T_n,\mathbf{x}^0)=\frac{k+1}{n}.$$

Since L-estimators are nondecreasing and translation equivariant, this is in obvious agreement with Theorem 2.3.1.

3.2.1 Sample median

Let us now consider some special types of L-estimators. The first one is the sample median which is defined as L-estimator such that $c_i = c_{n-i+1} = 0$ for $i = 0, 1, ..., k$ where k equals $\frac{n-2}{2}$ for n even and $\frac{n-1}{2}$ for n odd. From Theorem 3.2.1 if then follows next corollary.

Corollary 3.2.1. Let T_n be the sample median corresponding to $X_1, ..., X_n$. Then it holds that

$$\frac{n}{2} \leq \lim_{a \to \infty} \inf B(T_n, a) \leq \lim_{a \to \infty} \sup B(T_n, a) \leq \frac{n}{2} + 1$$

for n even and

$$\lim_{a \to \infty} B(T_n, a) = \frac{n+1}{2}$$

for n odd.

For the breakdown point of sample median it obviously holds that

$$\epsilon^*(T_n, \mathbf{x}^0) = \frac{1}{2}, \quad n \text{ even},$$
$$\epsilon^*(T_n, \mathbf{x}^0) = \frac{n+1}{2n}, \quad n \text{ odd},$$

Moreover the asymptotical breakdown point is

$$\epsilon^* = \lim_{n \to \infty} \epsilon^*(T_n, \mathbf{x}^0) = \frac{1}{2}.$$

It means that at least half of a random sample has to be replaced with infinity values to be able to cause the sample median to fail.

We can see that the sample median is robust estimator since the speed of convergence of its distribution tails does not depend on whether the random sample comes from light tailed distribution or heavy tailed one. Moreover, when the initial distribution is heavy tailed its distribution tails behaviour is much better that the one of the sample mean. However, the same does not hold when the initial distribution is light tailed.

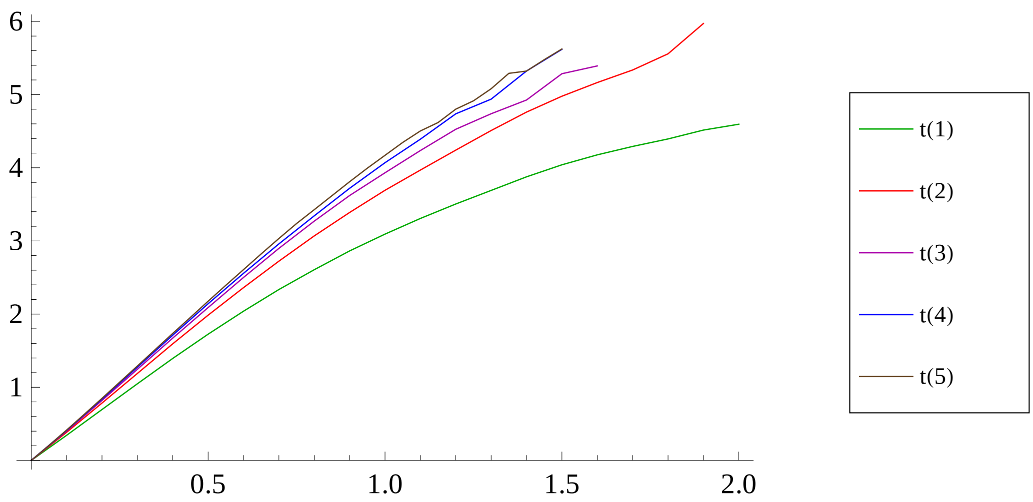
Similarly as for the sample mean we illustrate tail behavior of sample median for simulated random samples from Student t-distribution.

Illustration for Student t-distribution:

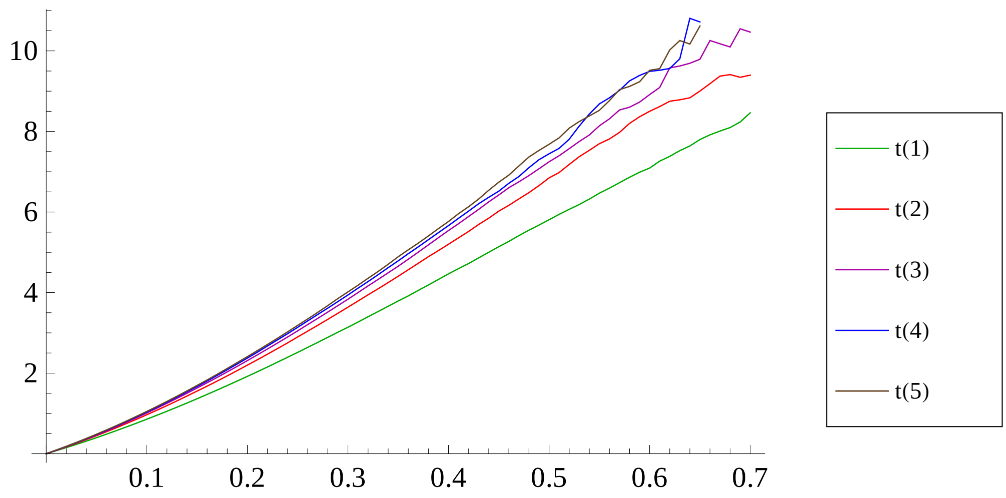
We have generated 10^6 of random samples from Student t-distribution with 1 to 5 degrees of freedom for $n = 20$ and $n = 100$ and calculated its corresponding sample medians. Since we chose n even, it holds that

$$10 \leq \liminf_{a\to\infty} B(T_n, a) \leq \limsup_{a\to\infty} B(T_n, a) \leq 11 \qquad \text{for } n = 20,$$
$$50 \leq \liminf_{a\to\infty} B(T_n, a) \leq \limsup_{a\to\infty} B(T_n, a) \leq 51 \qquad \text{for } n = 100.$$

Graphs 4 and 5 shows simulated values of $B(T_n, a)$ for $n = 20$ and $n = 100$.



Graph 4: Values of $B(a, T_n)$ for sample median and Student t-distribution with 1 to 5 degrees of freedom, $n = 20$.



Graph 5: Values of $B(a, T_n)$ for sample median and Student t-distribution with 1 to 5 degrees of freedom, $n = 100$.

It can be seen that the higher the degrees of freedom the t-distribution has the faster the convergence of $B(T_n, a)$ is. The slowest is for 1 degree of freedom. However for degrees of freedom higher than 2 the values of $B(T_n, a)$ are coming nearer hence the speed of the convergence is slowing down.

Tables 4 and 5 compares tails of standardized sample median T_n^s with tails of standardized normal distribution and t-distribution with $n-1$ degrees of freedom. Hence there are shown approximate values of $P_{t_k}(|T_n^s| > q)$ based on simulations where q are quantiles of normal or t-distribution with $n-1$ degrees of freedom.

$t_k \diagdown q$	$u_{1-\frac{\alpha}{2},\alpha=0.05}$	$t_{1-\frac{\alpha}{2},\alpha=0.05}$	$u_{1-\frac{\alpha}{2},\alpha=0.01}$	$t_{1-\frac{\alpha}{2},\alpha=0.01}$
t_1	0.053298	0.041456	0.016518	0.009770
t_2	0.052051	0.039155	0.013069	0.006609
t_3	0.051145	0.037960	0.011833	0.005661
t_4	0.051359	0.038001	0.011692	0.005351
t_5	0.050913	0.037466	0.010960	0.005091

Table 4: Simulated values of $P_{t_k}(|T_n^s| > q)$ for $n = 20$.[4]

[4] u_β is a β quantile of $N(0, 1)$ distribution and t_β is β quantile of Student t-distribution with $n-1$ degrees of freedom.

$t_k \diagdown q$	$u_{1-\frac{\alpha}{2},\alpha=0.05}$	$t_{1-\frac{\alpha}{2},\alpha=0.05}$	$u_{1-\frac{\alpha}{2},\alpha=0.01}$	$t_{1-\frac{\alpha}{2},\alpha=0.01}$
t_1	0.051110	0.037700	0.011421	0.005301
t_2	0.050437	0.036884	0.010650	0.004742
t_3	0.050353	0.036787	0.010503	0.004435
t_4	0.050346	0.036653	0.010320	0.004470
t_5	0.050289	0.036584	0.010307	0.004467

Table 5: Simulated values of $P_{t_k}(|T_n^s| > q)$ for $n = 100$.[5]

From Table 4 it can be seen that for $n = 20$ the tails of sample median are not far away from tails of normal distribution for $\alpha = 0.05$. However the more degrees of freedom the better the approximation is. For $\alpha = 0.01$ tails of sample median are quite well approximated by tails of normal distribution for degrees of freedom higher or equal to 3. For 1 degree of freedom are the tails of sample median better approximated by t-distribution with 19 degrees of freedom.

From Table 5 is can be seen that for $n = 100$ the tails of sample median are very similar to the one of normal distribution for $\alpha = 0.05$ and also for $\alpha = 0.01$.

Table 6 shows 0.975 quantiles of sample median based on simulations.

t_k	$q_{0.975}, n = 20$	$q_{0.975}, n = 100$
t_1	1.99405	1.97002
t_2	1.97934	1.96570
t_3	1.97265	1.96190
t_4	1.97128	1.96486
t_5	1.96746	1.96175

Table 6: Simulated values of 0.975 quantiles of sample median for $n = 20$ and $n = 100$ and t-distribution with k degrees of freedom.

[5] u_β is a β quantile of $N(0, 1)$ distribution and t_β is β quantile of Student t-distribution with $n - 1$ degrees of freedom.

3.2.2 Trimmed mean

Another special case of L-estimators is a trimmed mean which is of the form

$$T_n = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} X_i,$$

where $k < \frac{n}{2}, n \geq 3$.

Trimmed mean combines two types of creating estimators - trimming and averaging. Since under heavy tailed distribution the tail behaviour of sample mean is worse than the tail behaviour of sample median, we could expect the tail behaviour of trimmed mean to be worse than sample median but better then sample mean.

The tail behaviour of trimmed mean is described in the following theorem.

Theorem 3.2.2. Let $(X_1, \ldots, X_n)$ be a random sample from distribution with distribution function $F(x-\theta)$ and density $f(x-\theta)$ such that $f(-x) = f(x) > 0$. Let T_n, the estimator of θ, be the trimmed mean. Then if F has light tails then it holds that

$$n - 2k \leq \lim_{a\to\infty} \inf B(T_n, a) \leq \lim_{a\to\infty} \sup B(T_n, a) \leq n - k.$$

If $k < \frac{n-1}{2}$ and F has heavy tails then it holds

$$\lim_{a\to\infty} B(T_n, a) = k + 1.$$

Proof. We use again Lemma 3.1.1. Assume $\epsilon \in (0,1)$ and put $d_n = (1-\epsilon)(n-2k)$. Then it holds that

$$\begin{aligned}
E_0[\exp{(1-\epsilon)(n-2k)b|T_n|^r}] &= E_0\left[\exp\left((1-\epsilon)(n-2k)^{1-r}b|\sum_{i=k+1}^{n-k} X^{(i)}|^r\right)\right] \\
&\leq E_0\left[\exp\left((1-\epsilon)b\sum_{i=k+1}^{n-k} |X^{(i)}|^r\right)\right] \leq E_0\left[\exp\left((1-\epsilon)b\sum_{i=1}^{n} |X_i|^r\right)\right] \\
&\leq (E_0\left[\exp{((1-\epsilon)b|X_1|^r)}\right])^n < \infty
\end{aligned}$$

From Lemma 3.1.1 it follows that

$$\lim_{n\to\infty} \inf \left[\frac{-\log P_0(|T_n| > a)}{-\log(1-F(a))}\right] = \lim_{n\to\infty} \inf \left[\frac{-\log P_0(|T_n| > a)}{ba^r}\right] \geq (1-\epsilon)(n-2k)$$

for every $\epsilon \in (0,1)$ and it implies the first equality in Theorem.

Furthermore, it holds that

$$
\begin{aligned}
P_0(|T_n| > a) &= P_0(T_n > a) + P_0(T_n < -a) \\
&\geq P_0(X^{k+1} > -a, \ldots, X^{n-k-1} > -a, X^{n-k} > (2(n-2k)-1)a)) \\
&\qquad + P_0(X^{k+1} < a, \ldots, X^{n-k-1} > -a, X^{n-k} < -(2(n-2k)-1)a)) \\
&\geq P_0(X_1 > -a, \ldots, X_{n-k-1} > -a, X_{n-k} > (2(n-2k)-1)a, \\
&\qquad\qquad \ldots, X_n > (2(n-2k)-1)a) \\
&\qquad + P_0(X_1 < a, \ldots, X_{n-k-1} < a, X_{n-k} < -(2(n-2k)-1)a, \\
&\qquad\qquad \ldots, X_n < -(2(n-2k)-1)a) \\
&= 2F(a)^{n-k-1}(1 - F(2(n-k)-1)a)^{k+1},
\end{aligned}
$$

hence it holds that

$$
-\log P_0(|T_n| > a) \leq -(n-k-1)\log F(a) - (k+1)\log(1 - F(2(n-k)-1)a)
$$

and thus

$$
\begin{aligned}
\frac{-\log P_0(|T_n| > a)}{-\log(1-F(a))} &\leq (n-k-1)\frac{-\log F(a)}{-\log(1-F(a))} \\
&\quad + (k+1)\frac{-\log\left(1 - F(2(n-k)-1)a\right)}{m\log(2(n-k)-1)a)}\,\frac{m\log(2(n-k)-1)a)}{-\log(1-F(a))}
\end{aligned}
$$

from where it follows that

$$
\lim_{n\to\infty} \sup B(T_n, a) \leq k+1
$$

and the equality follows from Theorem 3.2.1. $\square$

From Theorem 3.2.2 follows that if F has light tails then for trimmed mean it holds that the higher k is the closer is the limit of $B(T_n, a)$ to one and hence the worse tail behaviour a trimmed mean has. On the other hand if F has heavy tails then the higher is k the better the tail behaviour of trimmed mean is.

Illustration for Student t-distribution:

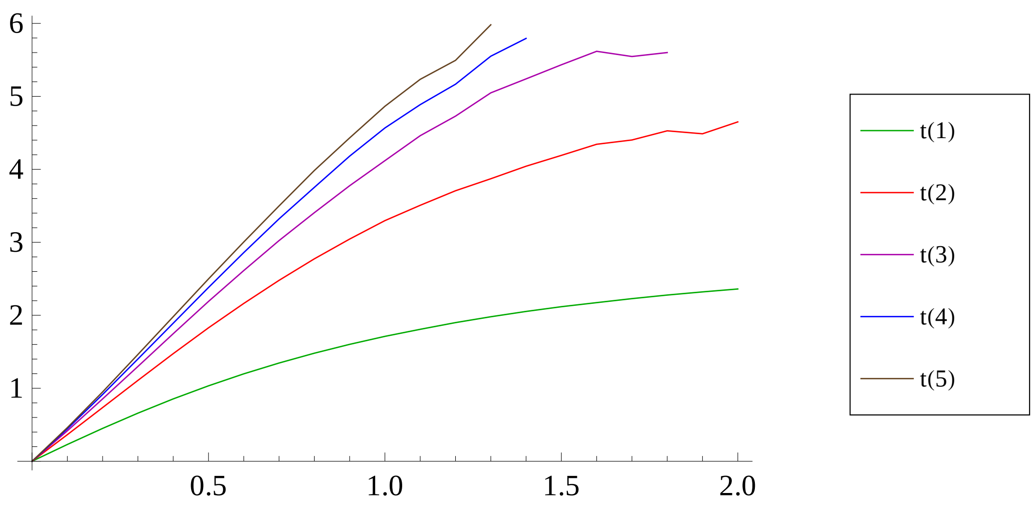
We have generated 10^6 of random samples $X_1, \ldots, X_n$ from Student t-distribution with 1 to 5 degrees of freedom for $n = 20$ and $n = 100$ and calculated its corresponding trimmed means where 0.1 of maximum observations and 0.1 of minimum observations were trimmed off. For $n = 20$ it means 2 maximum observations and 2 minimum observations trimmed off, hence $k = 2$. Since the t-distribution is heavy tailed from Theorem 3.2.2 it follows that

$$\lim_{a\to\infty} B(T_n, a) = 3.$$

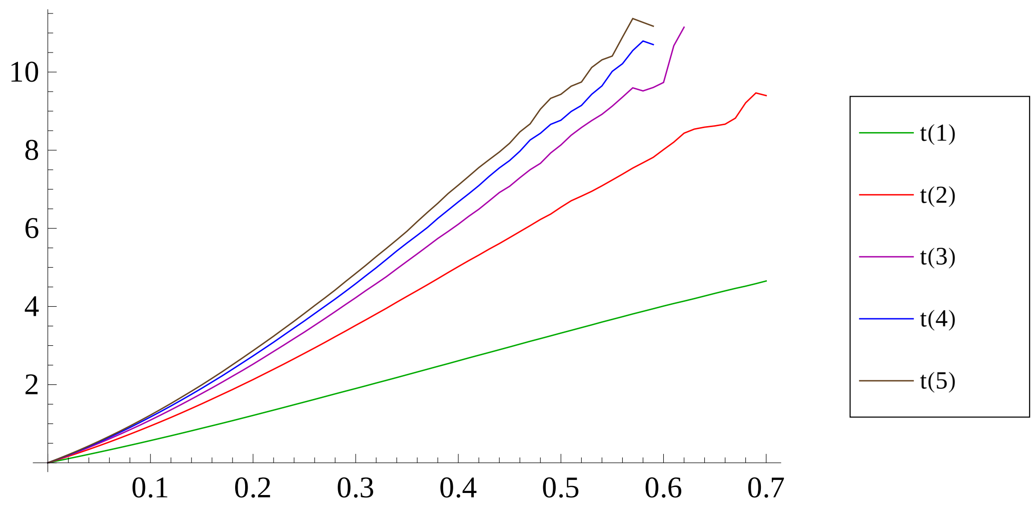
For $n = 100$ there were 10 maximum observations and 10 minimum observations trimmed off, hence $k = 10$ and from Theorem 3.2.2 it follows that

$$\lim_{a\to\infty} B(T_n, a) = 11.$$

Graphs 6 and 7 shows values of $B(T_n, a)$ for $n = 20$ and $n = 100$.



Graph 6: Values of $B(a, T_n)$ for trimmed mean and Student t-distribution with 1 to 5 degrees of freedom, $n = 20$.



Graph 7: Values of $B(a, T_n)$ for trimmed mean and Student t-distribution with 1 to 5 degrees of freedom, $n = 100$.

For $n = 20$ it can be seen that the behaviour of trimmed mean is similar to the sample mean in sense that although the limit of $B(T_n, a)$ is 3, for degrees of freedom higher than 1 values of $B(T_n, a)$ are higher than 3 for small $a > 0$ and then converges to 3. The more degrees of freedom the t-distribution has the higher values $B(T_n, a)$ reaches and the slower than the convergence is. For 1 degree of freedom this does not hold since values of $B(T_n, a)$ does not exceed 3.

For $n = 100$ it holds again that the more degrees of freedom the t-distribution has the higher values $B(T_n, a)$ has.

$t_k \diagdown q$	$u_{1-\frac{\alpha}{2},\alpha=0.05}$	$t_{1-\frac{\alpha}{2},\alpha=0.05}$	$u_{1-\frac{\alpha}{2},\alpha=0.01}$	$t_{1-\frac{\alpha}{2},\alpha=0.01}$
t_1	0.023755	0.017036	0.005114	0.002601
t_2	0.047455	0.035241	0.011173	0.005469
t_3	0.053468	0.039949	0.012802	0.006238
t_4	0.056871	0.042745	0.014046	0.006731
t_5	0.054962	0.043532	0.013900	0.006850

Table 7: Simulated values of $P_{t_k}(|T_n^s| > q)$ for trimmed mean, $n = 20$.[6]

[6] u_β is a β quantile of $N(0, 1)$ distribution and t_β is β quantile of Student t-distribution with $n - 1$ degrees of freedom.

$t_k \diagdown q$	$u_{1-\frac{\alpha}{2},\alpha=0.05}$	$t_{1-\frac{\alpha}{2},\alpha=0.05}$	$u_{1-\frac{\alpha}{2},\alpha=0.01}$	$t_{1-\frac{\alpha}{2},\alpha=0.01}$
t_1	0.037042	0.025960	0.006334	0.002534
t_2	0.049581	0.036153	0.0103072	0.004547
t_3	0.052695	0.038797	0.011465	0.004968
t_4	0.054037	0.039823	0.011838	0.005337
t_5	0.054761	0.040428	0.012162	0.005532

Table 8: Simulated values of $P_{t_k}(|T_n^s| > q)$ for trimmed mean, $n = 100$.[7]

Tables 7 and 8 compares tails of standardized trimmed mean T_n^s with tails of standardized normal distribution and t-distribution with $n-1$ degrees of freedom. There are shown approximate values of $P_{t_k}(|T_n^s| > q)$ based on simulations where q are quantiles of normal or t-distribution with $n-1$ degrees of freedom.

For $n = 20$ simulations shows that the tails of trimmed mean are closed to the tails of normal distribution for $\alpha = 0.05$ but not as closed as tails of sample mean. Furthermore, for 1 to 5 degrees of freedom, they are smaller than the tails of t-distribution with 19 degrees of freedom. For $\alpha = 0.01$ the tails of trimmed mean are quite good approximated for degrees of freedom higher or equal 2.

Table 9 shows 0.975 quantiles of standardized trimmed mean for $n = 20$ and $n = 100$.

t_k	$q_{0.975}, n = 20$	$q_{0.975}, n = 100$
t_1	1.96262	1.98051
t_2	1.98629	1.96455
t_3	1.97236	1.96713
t_4	1.96787	1.96023
t_5	1.96636	1.96246

Table 9: Simulated values of 0.975 quantiles of trimmed mean for $n = 20$ and $n = 100$.

[7] u_β is a β quantile of $N(0,1)$ distribution and t_β is β quantile of Student t-distribution with $n-1$ degrees of freedom.

3.3 M-estimators

M-estimators were first introduced by P. J. Huber in 1964. M-estimator is defined as a solution to minimization problem

$$\min_{\theta\in\Theta}\sum_{i=1}^{n}\rho(X_i,\theta),$$

where ρ is an arbitrary function. If ρ is differentiable in respect to θ with continuous derivative $\psi(X_i,\theta)=\frac{\partial\rho(X_i,\theta)}{\partial\theta}$ then M-estimator is a solution of equation

$$\sum_{i=1}^{n}\psi(X_i,\theta)=0,\quad \theta\in\Theta.$$

Remark 3.3.1. Simple example of M-estimator is obviously maximum likelihood estimator which is the solution of the equation

$$\sum_{i=1}^{n}(-\log f(X_i,\theta))=0,\quad \theta\in\Theta,$$

where f is corresponding density.

When θ denotes parameter of location than we usually suppose M-estimator to be a solution of

$$\min_{\theta\in\Theta}\sum_{i=1}^{n}\rho(X_i-\theta),$$

$$\text{or equivalently}\tag{3.3.1}$$

$$\sum_{i=1}^{n}\psi(X_i-\theta)=0,\quad \theta\in\Theta,$$

where ρ is chosen to be symmetric around 0. Uniqueness of solution of (3.3.1) depends on a form of function ρ respectively ψ. If ρ is linear and hence ψ is constant, there can be more than one solution of (3.3.1). Then usually one of these solution is chosen according to the rule

$$\begin{aligned}T_n&=\frac{1}{2}(T_n^+ + T_n^-),\ \text{where}\\ T_n^-&=\sup\{\theta:\sum_{i=1}^{n}\psi(X_i-\theta)>0\},\\ T_n^+&=\sup\{\theta:\sum_{i=1}^{n}\psi(X_i-\theta)<0\},\end{aligned}$$

Similar approach can be used when ψ is discontinuous function.

Also robustness of M-estimator depends on choice of ψ function. If ψ is unlimited than corresponding M-estimator is not robust. In such case it is enough one observation to be replaced by infinite value to cause the estimator to fail. Hence for the asymptotic breakdown point it holds that

$$\lim_{n\to\infty} \epsilon_n^* = 0.$$

Remark 3.3.2. Simple example of such M-estimator is sample mean which is a solution of

$$\sum_{i=1}^{n} \frac{1}{n}(X_i - t) = 0,$$

with respect to t, hence ψ is a linear unbounded function here.

On the other hand if ψ is bounded and odd then the corresponding M-estimator is robust and it holds the following theorem.

Theorem 3.3.1. Let $X_1, \ldots, X_n$ to be a random sample from distribution function $F(x - \theta)$ with the symmetric density $f(x - \theta)$. Let T_n be an M-estimator corresponding to the nondecreasing odd function ψ such that $\psi(x) = \psi(k)$ for $x \geq k, k > 0$. Then it holds that

$$\frac{n}{2} \leq \lim_{a\to\infty} \inf B(T_n, a) \leq \lim_{a\to\infty} \sup B(T_n, a) \leq \frac{n}{2} + 1$$

for n even and

$$\lim_{a\to\infty} B(T_n, a) = \frac{n+1}{2}$$

for n odd.

Proof. Firstly, suppose n even. Since T_n is obviously translation equivariant and ψ is odd nondecreasing function, it holds that

$$[T_n > a] = [T_n - a > 0] \supseteq \left[\sum_{i=1}^{n} \psi(X_i - a) > 0\right], \text{ and}$$

$$[T_n < -a] = [T_n + a < 0] \supseteq \left[\sum_{i=1}^{n} \psi(X_i + a) < 0\right],$$

From here it follows that

$$
\begin{aligned}
P_0(|T_n| > a) &= P_0(T_n > a) + P_0(T_n < -a) \\
&\geq P_0(\sum_{i=1}^{n} \psi(X_i - a) > 0) + P_0(\sum_{i=1}^{n} \psi(X_i + a) < 0) \\
&\geq P_0(X^{(\frac{n}{2})} - a > k) + P_0(X^{(\frac{n+2}{2})} + a < -k) \\
&= \frac{n}{2} \int_{k+a}^{\infty} \binom{n}{\frac{n}{2}} F(x)^{\frac{n}{2}-1} (1 - F(x))^{n-\frac{n}{2}} f(x) dx \\
&\qquad + \frac{n+2}{2} \int_{-\infty}^{-k-a} \binom{n}{\frac{n+2}{2}} F(x)^{\frac{n}{2}} (1 - F(x))^{n-\frac{n+2}{2}} f(x) dx \\
&= n \int_{0}^{1-F(k+a)} \binom{n}{\frac{n}{2}} (1-t)^{\frac{n}{2}-1} t^{\frac{n}{2}} dt \\
&\geq \frac{2n}{n+2} \binom{n}{\frac{n}{2}} F(k+a)^{\frac{n}{2}-1} (1 - F(k+a))^{\frac{n}{2}+1},
\end{aligned}
$$

where for the first integral substitution $1 - F(x) = t$ was used and for the second integral $F(x) = t$. Hence it holds that

$$
\frac{-\log P_0(|T_n| > a)}{-\log(1 - F(x))} \leq \left(\frac{n}{2} + 1\right) \frac{-\log(1 - F(k+a))}{-\log(1 - F(x))},
$$

hence

$$
\lim_{n\to\infty} \sup B(a, T_n) \leq \frac{n}{2} + 1.
$$

Similarly,

$$
\begin{aligned}
[T_n > a] = [T_n - a > 0] &\subseteq \left[\sum_{i=1}^{n} \psi(X_i - a) \geq 0\right], \text{ and} \\
[T_n < -a] = [T_n + a < 0] &\subseteq \left[\sum_{i=1}^{n} \psi(X_i + a) \leq 0\right],
\end{aligned}
$$

From here it follows that

$$
\begin{aligned}
P_0(|T_n| > a) &= P_0(T_n > a) + P_0(T_n < -a) \\
&\leq P_0\left(\sum_{i=1}^{n} \psi(X_i - a) \geq 0\right) + P_0\left(\sum_{i=1}^{n} \psi(X_i - a) \leq 0\right) \\
&\leq P_0(X^{(\frac{n}{2}+1)} - a \geq -k) + P_0(X^{(\frac{n}{2})} + a \leq k) \\
&= \frac{n+2}{2} \int_{-k+a}^{\infty} \binom{n}{\frac{n+2}{2}} F(x)^{\frac{n}{2}} (1 - F(x))^{n-\frac{n+2}{2}} f(x) dx \\
&\qquad + \frac{n}{2} \int_{-\infty}^{k-a} \binom{n}{\frac{n}{2}} F(x)^{\frac{n}{2}-1} (1 - F(x))^{n-\frac{n}{2}} f(x) dx \\
&= 2n \binom{n-1}{\frac{n}{2}} \int_{F(a-k)}^{1} t^{\frac{n}{2}} (1-t)^{\frac{n}{2}-1} dt \leq 4 \binom{n-1}{\frac{n}{2}} (1 - F(a-k))^{\frac{n}{2}},
\end{aligned}
$$

hence

$$\lim_{n\to\infty} \inf B(a, T_n) \geq \frac{n}{2}.$$

□

Theorem 3.3.1 tells that when ψ is chosen having some special properties then the tail behaviour of M-estimator is similar to the one of sample median. Such M-estimator has also same breakdown point as median, hence it holds

$$\epsilon^*(T_n, \mathbf{x}^0) = \frac{1}{2}, \quad n \text{ even},$$
$$\epsilon^*(T_n, \mathbf{x}^0) = \frac{n+1}{2n}, \quad n \text{ odd},$$

and the asymptotical breakdown point is

$$\epsilon^* = \lim_{n\to\infty} \epsilon^*(T_n, \mathbf{x}^0) = \frac{1}{2}.$$

Since robustness of M-estimator essentially depends on choice of function ψ, there appears a question according to which criteria to choose this function. P. J. Huber suggest function which is linear in bounded interval $[-k, k]$ and constant outside this interval. He also give reasons for such estimator to be convenient for neighborhood of normal distribution. This type of estimator is often called Huber's estimator in literature.

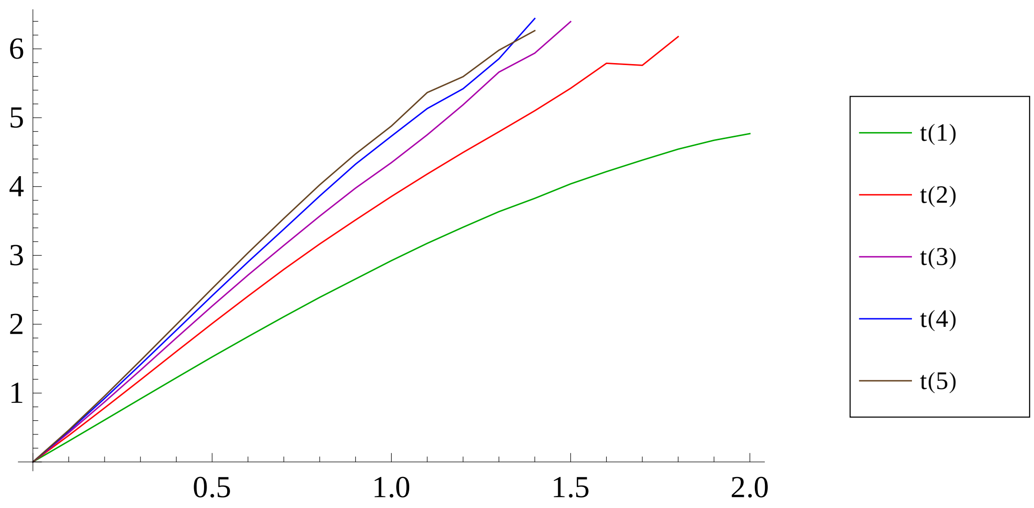
Illustration for Student t-distribution

We have generated 10^6 of random samples from Student t-distribution with 1 to 5 degrees of freedom for $n = 20$ and $n = 100$ and calculated its corresponding Huber estimators for ψ defined as

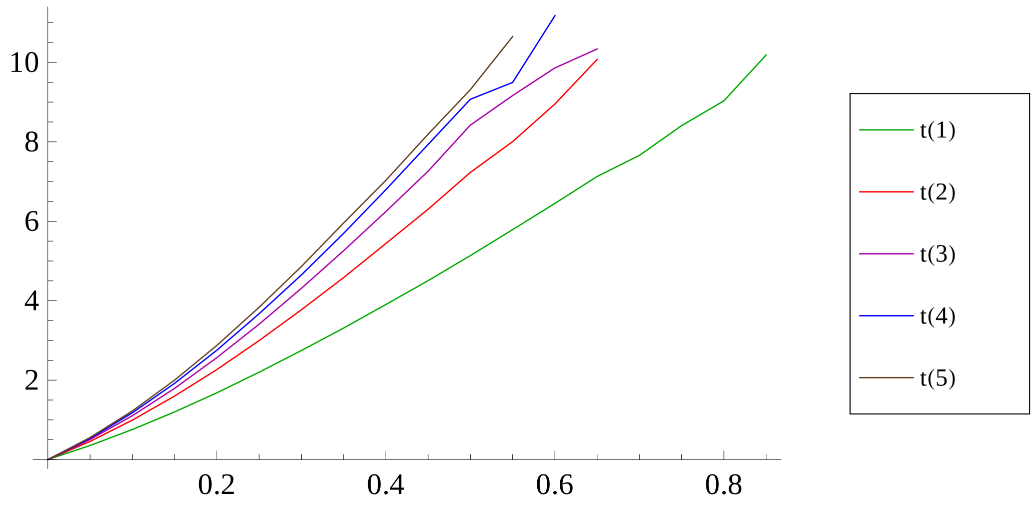
$$\psi(x) = \begin{cases} x & \text{for } |x| \leq 1.5 \\ 1.5 \operatorname{sgn}(x) & \text{for } |x| > 1.5 \end{cases}$$

From Theorem 3.3.1 it follows that bounds for convergence of $B(T_n, a)$ are the same as for the sample median, hence $[10, 11]$ for $n = 20$ and $[50, 51]$ for $n = 100$.

Graphs 8 and 9 shows values of $B(T_n, a)$ for standardized Huber estimator and for $n = 20$ and $n = 100$.



Graph 8: Values of $B(a, T_n)$ for Huber estimator and Student t-distribution with 1 to 5 degrees of freedom, $n = 20$.



Graph 9: Values of $B(a, T_n)$ for Huber estimator and Student t-distribution with 1 to 5 degrees of freedom, $n = 100$.

The tail behaviour of standardized Huber estimator is similar to sample median in sense that the higher degrees of freedom t-distribution has the higher values $B(T_n, a)$ reaches and the fastest the convergence is for small a.

Tables 10 and 11 compares tails of standardized trimmed mean T_n^s with tails of standardized normal distribution and t-distribution with $n-1$ degrees of freedom. There are shown approximate values of $P_{t_k}(|T_n^s| > q)$ based on simulations where q are quantiles of normal or t-distribution with $n-1$ degrees of freedom.

$t_k \diagdown q$	$u_{1-\frac{\alpha}{2},\alpha=0.05}$	$t_{1-\frac{\alpha}{2},\alpha=0.05}$	$u_{1-\frac{\alpha}{2},\alpha=0.01}$	$t_{1-\frac{\alpha}{2},\alpha=0.01}$
t_1	0.052068	0.039551	0.013972	0.007437
t_2	0.051189	0.038041	0.012155	0.005904
t_3	0.050917	0.037793	0.011630	0.005384
t_4	0.050860	0.037333	0.011074	0.004956
t_5	0.050601	0.037207	0.010932	0.004928

Table 10: Simulated values of $P_{t_k}(|T_n^s| > q)$ for Huber estimator, $n = 20$.[8]

$t_k \diagdown q$	$u_{1-\frac{\alpha}{2},\alpha=0.05}$	$t_{1-\frac{\alpha}{2},\alpha=0.05}$	$u_{1-\frac{\alpha}{2},\alpha=0.01}$	$t_{1-\frac{\alpha}{2},\alpha=0.01}$
t_1	0.050706	0.037144	0.010912	0.004821
t_2	0.050350	0.036806	0.010519	0.004524
t_3	0.050035	0.036524	0.010183	0.004440
t_4	0.049978	0.036387	0.010230	0.004353
t_5	0.050250	0.036607	0.010195	0.004378

Table 11: Simulated values of $P_{t_k}(|T_n^s| > q)$ for Huber estimator, $n = 100$.

Tables 10 and 11 gives similar conclusions as in the case of sample median. For $n = 20$ the more degrees of freedom the better the approximation of the tails of Huber estimator by the tails of normal distribution is for $\alpha = 0.05$, The same holds for $\alpha = 0.01$, there are the tails quite good approximated for degrees of freedom higher than 3. For $n = 100$ the tails of Huber estimator for all degrees of freedom are well approximated by the tails of normal distribution for both $\alpha = 0.05$ and $\alpha = 0.01$.

Table 12 shows 0.975 quantiles of standardized Huber estimator for $n = 20$ and $n = 100$.

[8] u_β is a β quantile of $N(0,1)$ distribution and t_β is β quantile of Student t-distribution with $n-1$ degrees of freedom.

t_k	$q_{0.975}, n = 20$	$q_{0.975}, n = 100$
t_1	1.97866	1.96545
t_2	1.97041	1.95821
t_3	1.96956	1.95980
t_4	1.96811	1.95891
t_5	1.96660	1.96277

Table 12: Simulated values of 0.975 quantiles of Huber estimator for $n = 20$ and $n = 100$.

3.4 R-estimators

Tail behavior of R-estimators was shown by Jurečková [1981] and moreover by Zuo [2000].

Family of R-estimators was introduced by Hodges and Lehmann in 1963. It is based on rank tests. Suppose $X_1, \dots, X_n$ to be a random sample from distribution function $F(x-\theta)$ with symmetric density $f(x-\theta)$. Denote $R_i^+(\theta)$ rank of $|X_i - \theta|$ among $|X_1 - \theta|, \dots, |X_n - \theta|$, hence

$$R_i^+(\theta) = \sum_{j=1}^{n} \mathbb{I}_{[|X_j-\theta| \le |X_i-\theta|]}, \quad i = 1, \dots, n.$$

Then for testing hypothesis about center of symmetry θ

$$H_0 : \theta = \theta_0$$

signed-rank test can be used based on statistics

$$S_n(\theta_0) = \sum_{i=1}^{n} \operatorname{sgn}(X_i - \theta_0) a_n(R_i^+(\theta)),$$

where $a_n(1) \le \dots \le a_n(n)$ are given scores.

As an R-estimator was suggested value of t such that $S_n(t) = 0$. Since such solution does not have to exist, an R-estimator is generally defined as

$$T_n = \frac{1}{2}(T_n^+ + T_n^-), \text{ where}$$

$$T_n^- = \sup\{t : S_n(t) > 0\},$$

$$T_n^+ = \sup\{t : S_n(t) < 0\}.$$

Remark 3.4.1. If we put $a_n(i) = 1$ for $i = 1, \dots, n$, we get T_n as a solution of equation

$$\sum_{i=1}^{n} \operatorname{sgn}(X_i - t) = 0$$

with respect to t which is satisfied only for middle value of a sample, hence for sample median. Sample median is thus a special case of R-estimators.

If we put $a_n(i) = \frac{i}{n+1}$ for $i = 1, \dots, n$, then $S_n(\theta_0)$ is a statistics of One-sample Wilcoxon signed rank test. Estimator based on this statistics is called Hodges-Lehmann estimator and can be easily computed as median of means of all pairs of observations. Hence Hodges-Lehmann estimator can be written as

$$T_n = \operatorname{med} \left\{ \frac{X_i + X_j}{2}; 1 \leq i \leq j \leq n \right\}.$$

Tail behaviour of Hodges Lehmann estimator is described in next Theorem.

Theorem 3.4.1. Let $X_1, \dots, X_n$ be a random sample from distribution function $F(x - \theta)$ with symmetric density $f(x - \theta)$. Let T_n be the Hodges Lehmann estimator. Then it holds that

$$k_n + 1 \leq \lim_{a \to \infty} \inf B(a, T_n) \leq \lim_{a \to \infty} \sup B(a, T_n) \leq n - k_n,$$

where $k_n = \max\{m \in \mathbb{N} : m \leq 0.2n\}$.

For proof of Theorem 3.4.1 following Lemma will be used.

Lemma 3.4.1. Let $y_1, \dots, y_n$ be integers such that $|y_i| = i$ for $i = 1, \dots, n$. If at least $0.8n$ those integers are negative, then $\sum_{i=1}^{n} y_i < 0$.

Proof of Lemma 3.4.1. It holds that

$$\sum_{i=1}^{n} y_i \leq -\sum_{i=1}^{0.8n} i + \sum_{i=0.8n+1}^{n} i < 0.$$

for both $0.8n$ integer and also if $0.8n$ is not an integer.

$\square$

Proof of Theorem 3.4.1. From definition of Hodges Lehmann estimator it holds that

$$\begin{aligned}
T_n &= \frac{1}{2}(T_n^+ + T_n^-), \text{ where} \\
T_n^- &= \sup\{t : \sum_{i=1}^{n} \operatorname{sgn}(X_i - t)R_i^+(t) > 0\}, \\
T_n^+ &= \inf\{t : \sum_{i=1}^{n} \operatorname{sgn}(X_i - t)R_i^+(t) < 0\}.
\end{aligned}$$

From here and the fact that R-estimator is translation equivariant it follows that

$$\begin{aligned}
P_0(|T_n| > a) &\leq 2P_0(\sum_{i=1}^{n} \operatorname{sgn}(X_i - a)R_i^+(a) \geq 0) \leq 2P_0(X^{(n-k_n)} \geq a) \\
&= 2n\binom{n-1}{k_n}\int_a^{\infty}(1-F(x))^{n-n+k_n}F(x)^{n-k_n-1}f(x)dx \\
&= 2n\binom{n-1}{k_n}\int_0^{1-F(a)}(t)^{k_n}(1-t)^{n-k_n-1}dt \\
&\leq 2n\binom{n-1}{k_n}\frac{1-F(a)^{k_n+1}}{k_n+1},
\end{aligned}$$

where for the integral substitution $1 - F(x) = t$ was used. From here it follows that

$$\lim_{n\to\infty} \inf B(a, T_n) \geq k_n + 1.$$

Furthermore,

$$\begin{aligned}
P_0(|T_n| > a) &\geq 2P_0(\sum_{i=1}^{n} \operatorname{sgn}(X_i - a)R_i^+(a) > 0) \geq 2P_0(X^{(k_n+1)} > a) \\
&= 2n\binom{n-1}{k_n}\int_a^{\infty}(1-F(x))^{n-k_n-1}F(x)^{k_n}f(x)dx \\
&= 2n\binom{n-1}{k_n}\int_0^{1-F(a)}t^{n-k_n-1}(1-t)^{k_n}dt \\
&\geq 2n\binom{n-1}{k_n}1-F(a)^{k_n}\int_0^{1-F(a)}t^{n-k_n-1}dt \\
&\geq 2n\binom{n-1}{k_n}(1-F(a))^{k_n}\frac{(1-F(a))^{n-k_n}}{n-k_n},
\end{aligned}$$

where for the integral was used same substitution $1 - F(x) = t$. This gives

$$\lim_{n\to\infty} \sup B(a, T_n) \leq n - k_n.$$

□

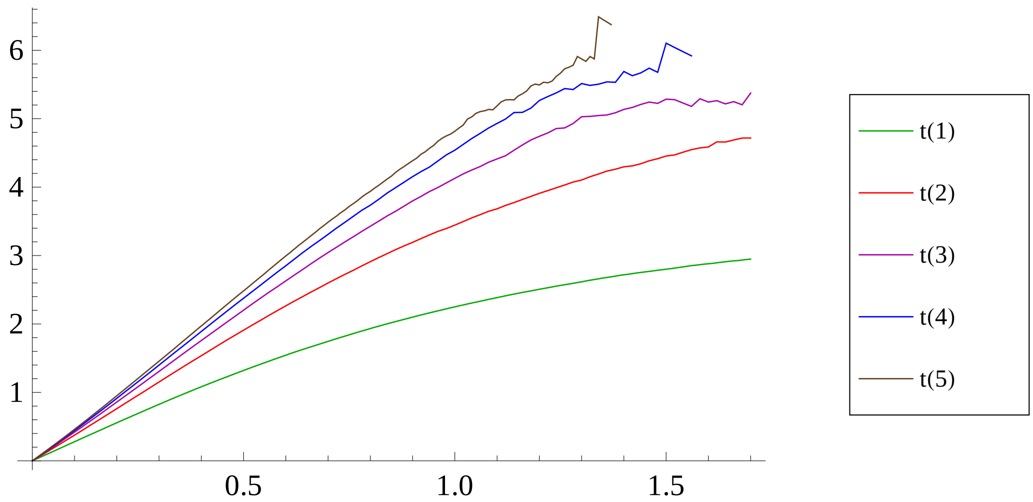
Theorem 3.4.1 tells that the higher the sample size is the higher the lower and the upper bounds for convergence of $B(T_n, a)$ are. Moreover, the wider range of the bounds is. From here it follows that for the bigger sample size the wider values the limit of $B(T_n, a)$ can have. We will illustrate tail behaviour of Hodges Lehmann estimator for t-distribution with 1 to 5 degrees of freedom.

Illustration for Student t-distribution

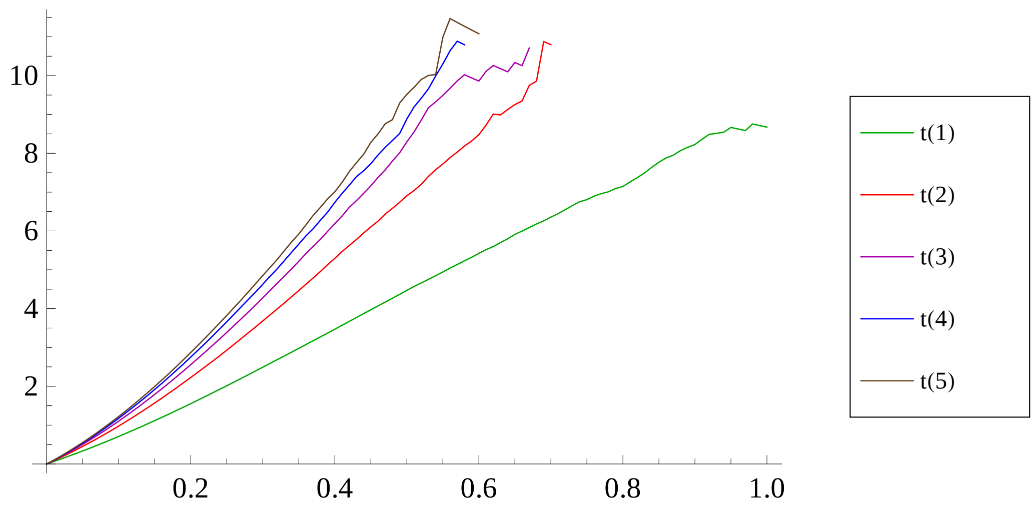
We have generated 10^6 of random samples from Student t-distribution with 1 to 5 degrees of freedom for $n = 20$ and $n = 100$ and calculated its corresponding Hodges Lehmann estimators. From Theorem 3.4.1 it holds that

$$5 \leq \lim_{a\to\infty} \inf B(T_n, a) \leq \lim_{a\to\infty} \sup B(T_n, a) \leq 16 \quad \text{for } n = 20,$$
$$21 \leq \lim_{a\to\infty} \inf B(T_n, a) \leq \lim_{a\to\infty} \sup B(T_n, a) \leq 80 \quad \text{for } n = 100.$$

Graphs 10 and 11 shows values of $B(T_n, a)$ for $n = 20$ and $n = 100$.



Graph 10: Values of $B(a, T_n)$ for Hodges Lehmann estimator and Student t-distribution with 1 to 5 degrees of freedom, $n = 20$.



Graph 11: Values of $B(a, T_n)$ for Hodges Lehmann estimator and Student t-distribution with 1 to 5 degrees of freedom, $n = 100$.

From Graph 10 it can be seen that for $n = 20$ the more degrees of freedom the t-distribution has the higher values $B(T_n, a)$ reaches. Moreover, for a lower number of degrees of freedom $B(T_n, a)$ converges rather to lower values than for more degrees of freedom.

Tables 13 and 14 compares tails of standardized Hodges Lehmann estimator T_n^s with tails of standardized normal distribution and t-distribution with $n-1$ degrees of freedom. There are shown approximate values of $P_{t_k}(|T_n^s| > q)$ based on simulations where q are quantiles of normal or t-distribution with $n-1$ degrees of freedom.

$t_k \diagdown q$	$u_{1-\frac{\alpha}{2},\alpha=0.05}$	$t_{1-\frac{\alpha}{2},\alpha=0.05}$	$u_{1-\frac{\alpha}{2},\alpha=0.01}$	$t_{1-\frac{\alpha}{2},\alpha=0.01}$
t_1	0.052262	0.041934	0.019407	0.012588
t_2	0.052163	0.039479	0.013670	0.007203
t_3	0.051302	0.038207	0.012073	0.005792
t_4	0.050990	0.037584	0.011353	0.005265
t_5	0.050964	0.037476	0.010961	0.005025

Table 13: Simulated values of $P_{t_k}(|T_n^s| > q)$ for Hodges Lehman estimator, $n = 20$.[9]

[9] u_β is a β quantile of $N(0,1)$ distribution and t_β is β quantile of Student t-distribution with

$t_k \diagdown q$	$u_{1-\frac{\alpha}{2},\alpha=0.05}$	$t_{1-\frac{\alpha}{2},\alpha=0.05}$	$u_{1-\frac{\alpha}{2},\alpha=0.01}$	$t_{1-\frac{\alpha}{2},\alpha=0.01}$
t_1	0.051479	0.038523	0.012224	0.005998
t_2	0.050491	0.036999	0.010680	0.004601
t_3	0.050417	0.036761	0.010342	0.004468
t_4	0.050110	0.036608	0.010321	0.004481
t_5	0.050054	0.036547	0.010226	0.004375

Table 14: Simulated values of $P_{t_k}(|T_n^s| > q)$ for Hodges Lehmann estimator, $n = 100$.[10]

For both $n = 20$ and $n = 100$ the tails of Hodges Lehmann estimator are quite good approximated by the tails of normal distribution for $\alpha = 0.05$ and the more degrees of freedom the better the approximation is. The similar conclusion holds for $\alpha = 0.01$.

Table 15 shows 0.975 quantiles of standardized Hodges Lehmann estimator for $n = 20$ and $n = 100$.

t_k	$q_{0.975}, n = 20$	$q_{0.975}, n = 100$
t_1	1.98501	1.96907
t_2	1.98191	1.96630
t_3	1.97348	1.96189
t_4	1.96999	1.96000
t_5	1.96824	1.95877

Table 15: Simulated values of 0.975 quantiles of Hodges Lehmann estimator for $n = 20$ and $n = 100$.

$n - 1$ degrees of freedom.

[10] u_β is a β quantile of $N(0, 1)$ distribution and t_β is β quantile of Student t-distribution with $n - 1$ degrees of freedom.

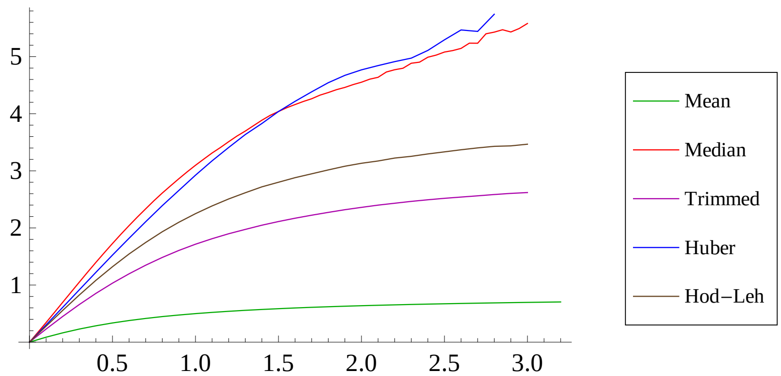
3.5 Comparison of estimators tail behaviour

In this section we compare tail behaviour of above mentioned estimators with each other for t-distribution with 1 to 5 degrees of freedom. Table 16 shows lower and upper bound for $B(T_n, a)$ for $n = 20$ as follows from Theorems described in previous sections. Moreover simulated variances of estimators are included.

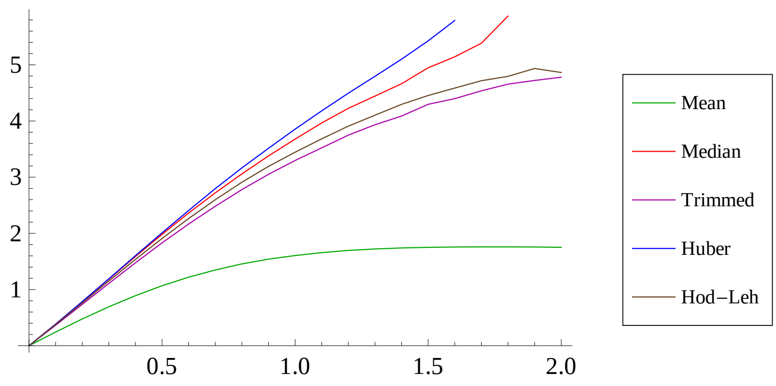
T_n	L	U	var_{t_1}	var_{t_2}	var_{t_3}	var_{t_4}	var_{t_5}
Mean	1	1	$5.9*10^5$	0.782	0.152	0.100	0.083
Median	10	11	0.140	0.101	0.091	0.086	0.083
Trimmed Mean	3	3	0.413	0.114	0.085	0.074	0.069
Huber	10	11	0.163	0.099	0.081	0.073	0.068
Hodges Lehmann	5	16	0.235	0.107	0.084	0.075	0.069

Table 16: L, U - lower and upper bound for limit of $B(T_n, a)$, var_{t_k} - simulated variance of T_n for t-distribution with k degrees of freedom, $n = 20$.

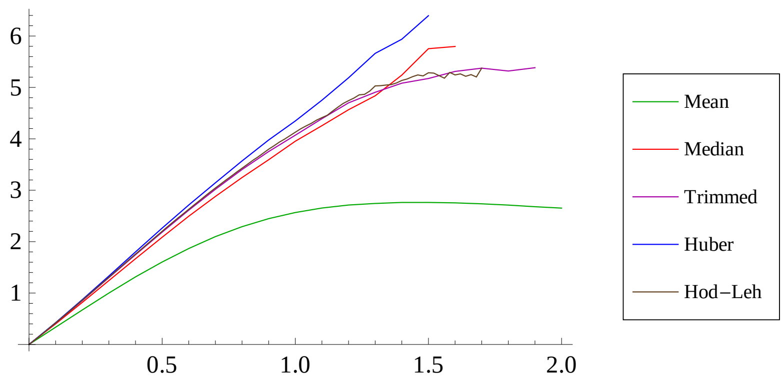
Graphs 12 - 16 shows simulated values of $B(T_n, a)$ compared for mentioned estimators for $n = 20$.



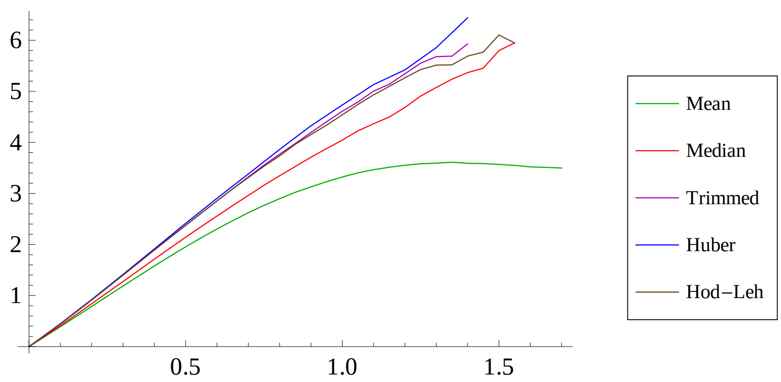
Graph 12: Values of $B(a, T_n)$ for Student t-distribution with 1 degree of freedom, $n = 20$.



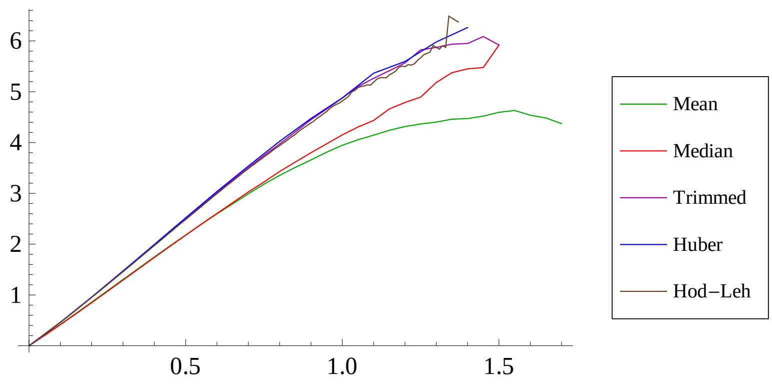
Graph 13: Values of $B(a, T_n)$ for Student t-distribution with 2 degrees of freedom, $n = 20$.



Graph 14: Values of $B(a, T_n)$ for Student t-distribution with 3 degrees of freedom, $n = 20$.



Graph 15: Values of $B(a, T_n)$ for Student t-distribution with 4 degrees of freedom, $n = 20$.



Graph 16: Values of $B(a, T_n)$ for Student t-distribution with 5 degrees of freedom, $n = 20$.

Simulations compares speed of the convergence of tails of estimators. For 1 degree of freedom we can see that $B(T_n, a)$ of mean and trimmed mean is converging to their limits from the bottom and hence they are estimators with the slowest converging tails. Similarly Hodges-Lehman estimator is converging very slowly and simulated values of $B(T_n, a)$ are below lower bound of its limit. The fastest converging tails has sample median and Huber estimator in this case.

For 2 degrees of freedom the speed of convergence of tails of median, trimmed mean, Huber and Hodges Lehmann estimator is getting nearer. For 3 degrees of freedom the speed of convergence of tails of these estimators is comparable. The fastest tails in these cases showed Huber estimator.

For 4 degrees of freedom the speed of converging tails of trimmed mean, Huber estimator and Hodges-Lehmann estimator exceeds the speed of tails of median. Moreover, the speed of convergence of tails of sample mean is getting nearer to the ones of other estimators for small a.

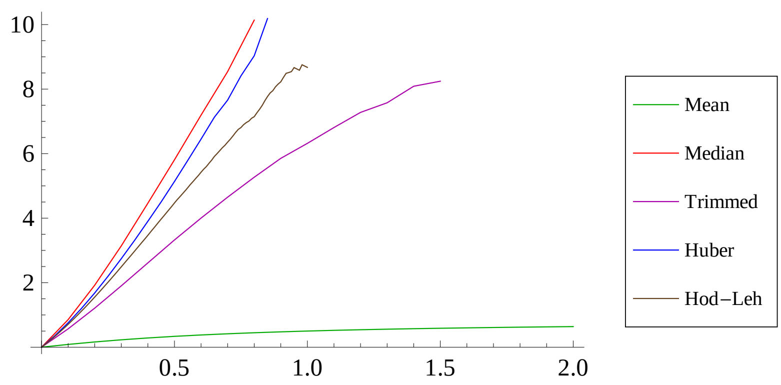
For 5 degrees of freedom the fastest tails has trimmed mean, Huber estimator and Hodges Lehmann estimator, the slowest tails has median and the sample mean which have similar convergence of tails for small a, for higher a the tails of median are converging faster.

Hence according to simulations the estimators with the best tail behaviour is median and Huber estimator for 1 degree of freedom. With more degrees of freedom estimators with best tail behaviour are becoming trimmed mean, Huber estimator and Hodges Lehmann estimator. However, our conclusions based on simulations are made only for limited a since it is very difficult to make simulations for extreme values of a. Convergence of very extreme tails can then be judge according to bounds of limit of $B(T_n, a)$, hence trimmed mean and sample mean are the estimators with poorest tail behaviour for very extreme values of a.

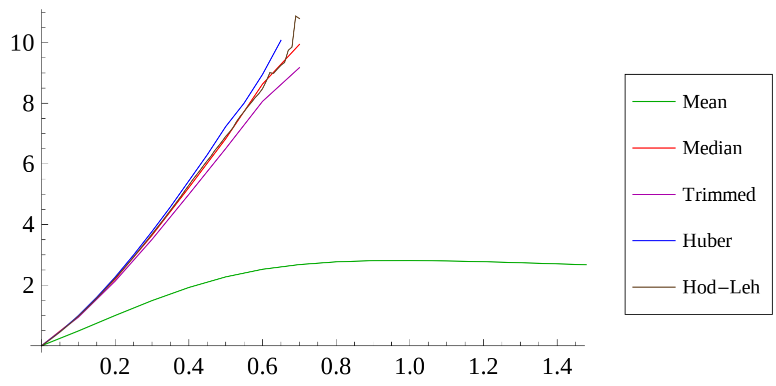
Table 17 shows lower and upper bound for $B(T_n, a)$ and simulated variances of estimators for $n = 100$. Graphs 17 - 21 shows simulated values of $B(T_n, a)$ compared for mentioned estimators for $n = 100$ .The conclusions are similar to the case with the sample size 20.

	L	U	var_{t_1}	var_{t_2}	var_{t_3}	var_{t_4}	var_{t_5}
Mean	1	1	$2.8*10^9$	0.296	0.030	0.020	0.017
Median	50	51	0.025	0.020	0.018	0.018	0.017
Trimmed Mean	11	11	0.052	0.021	0.016	0.015	0.014
Huber	50	51	0.030	0.019	0.016	0.014	0.014
Hodges Lehmann	21	80	0.035	0.020	0.016	0.014	0.014

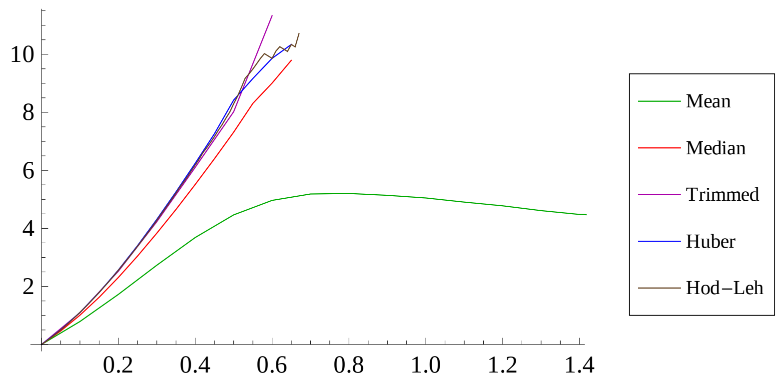
Table 17: L, U - lower and upper bound for limit of $B(T_n, a)$, var_{t_k} - simulated variance for t-distribution with k degrees of freedom, $n = 100$.



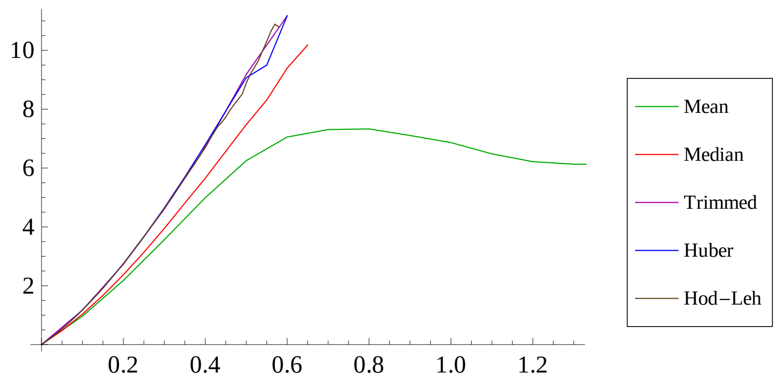
Graph 17: Values of $B(a, T_n)$ for Student t-distribution with 1 degree of freedom, $n = 20$.



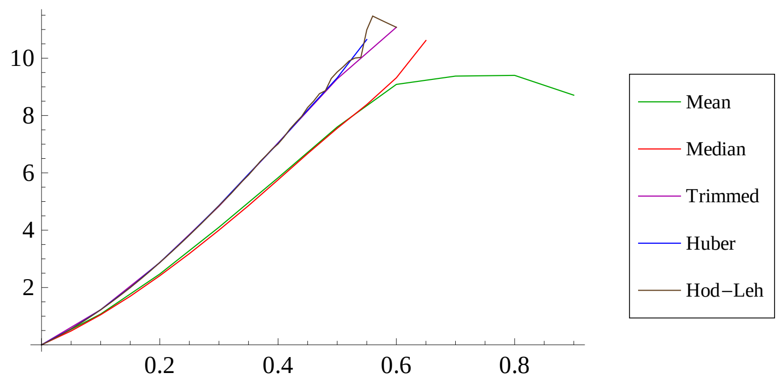
Graph 18: Values of $B(a, T_n)$ for Student t-distribution with 2 degrees of freedom, $n = 20$.



Graph 19: Values of $B(a, T_n)$ for Student t-distribution with 3 degrees of freedom, $n = 20$.



Graph 20: Values of $B(a, T_n)$ for Student t-distribution with 4 degrees of freedom, $n = 20$.



Graph 21: Values of $B(a, T_n)$ for Student t-distribution with 5 degrees of freedom, $n = 20$.

Chapter 4

Distribution of estimators

Asymptotic and finite sample distributions of robust estimators are widely discussed by Jurečková [2001], [2012] or by Jurečková and Picek [2011]. In the next sections we will shows the basic idea of how it can be shown that the asymptotic distribution of before mentioned estimators is normal, hence light tailed one, however, for any finite sample size the distribution of such estimator is heavy tailed whenever the random sample comes from heavy tailed distribution.

4.1 Asymptotic distribution

Definition 8. We say that the functional T is differentiable within the *Gâteaux meaning* according to P in the direction of Q, if the limit

$$T'_Q(P) = \lim_{t \to 0_+} \frac{T(P + t(Q - P)) - T(P)}{t}$$

exists. $T'_Q(P)$ is called *Gâteaux derivation* according to P in the direction of Q.

Definition 9. We say that the functional T is differentiable within the *Fréchet meaning* according to P if there exists linear functional $L_P(Q - P)$ such that for $Q \in \mathcal{P}, d(P, Q) \leq C$ for any fixed $C \in (0, \infty)$

$$\lim_{t \to 0} \frac{T(P + t(Q - P)) - T(P)}{t} = L_P(Q - P)$$

uniformly. Functional $L_P(Q - P)$ is called *Fréchet derivation* according to P in the direction of Q.

Remark 4.1.1. If P_n is an empirical distribution of random vector $(X_1, \dots, X_n)$ then it holds that

$$L_P(P_n - P) = T'_{P_n}(P).$$

Definition 10. As *an influence function* of a functional T in probability distribution P is understand a Gâteaux derivation T according to P in the direction of δ_x. We will denote

$$IF(x; T, P) = T'_x(P).$$

Theorem 4.1.1. Let T be a functional, differentiable within the Fréchet meaning according to P and suppose that empirical distribution P_n of a random vector $(X_1, \dots, X_n)$ satisfies

$$\sqrt{n} d(P_n, P) = O_P(1) \tag{4.1.1}$$

for $n \to \infty$. If Gâteaux derivation $T'_{X_1}(P)$ have positive variance $\text{var}_P T'_{X_1}(P) > 0$, then sequence $\sqrt{n}(T(P_n) - T(P))$ has asymptotic normal distribution with zero mean and variance $\text{var}_P T'_{X_1}(P)$.

Remark 4.1.2. It holds that

$$\sqrt{n}(T(P_n) - T(P)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} T'_{X_i}(P) + o_P(1).$$

If T_n is differentiable within the Fréchet meaning and (4.1.1) holds then

$$\text{var}_P IF(X_1; T, P) = E_P(IF(X_1; T, P))^2 > 0$$

. Hence if for estimator T_n holds

$$\sqrt{n}(T_n - T(F)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF(X_i; T, F) + o_P(1)$$

then T_n has asymptotic normal distribution with variance $E_F(IF(X_1; T, F))^2$.

This conclusion can then be applied on M, L and R-estimators which influence functions are derived for example by Jurečková [2001] and got following results (the same terms are used as in previous chapter):

- *M-estimators.* Let $\rho(x)$ be an absolutely continuous function with derivation $\psi(x)$ and let function $h(t) = \int_{\mathbb{R}} \rho(x - t) dF(x)$ has unique minimum in

$t = 0$. If ψ is absolutely continuous with derivation ψ' and $\gamma = \int \psi'(x)dF(x) > 0$ then there exists sequence T_n of roots of equation $\sum_{i=1}^n \psi(X_i - t) = 0$ such that for $n \to \infty$

$$P_\theta(\sqrt{n}(T_n - \theta) \leq x) \to \Phi\left(0, \frac{x}{\sigma^2(\psi, F)}\right),$$

where $\sigma^2(\psi, F) = \gamma^{-2} \int_{\mathbb{R}} \psi^2(x)dF(x)$ and Φ is a distribution function of $N(0,1)$.

- *L-estimators.* Suppose F is continuous almost everywhere and F^{-1} is lipschitz in neighborhood of points of discontinuity of weight function J, which determines weights of L-estimator as $c_i = \int_{\frac{i-1}{n}}^{\frac{i}{n}} J(s)ds$. Let $T(F) = \int_0^1 J(u)F^{-1}(u)du$. Then for $n \to \infty$

$$P_\theta(\sqrt{n}(T_n - T(F)) \leq x) \to \Phi\left(0, \sigma^2(J, F)\right),$$

where $\sigma^2(J, F) = \int_{\mathbb{R}} \left[\int_R (\mathbb{I}_{[y \geq x]} - F(y))J(F(y))dy\right]^2 dF(x)$.

- *R-estimators.* Let scores of an R-estimator $a_n(i)$ be generated by nondecreasing score function $\varphi : [0, 1) \mapsto \mathbb{R}$, $\varphi(0) = 0$ such that $a_n(i) = \varphi\left(\frac{i}{n+1}\right)$. Suppose F having an absolute continuous density f and finite Fisher information $\mathcal{I}(F)$. Then for $n \to \infty$

$$P_\theta(\sqrt{n}(T_n - \theta) \leq x) \to \Phi\left(0, \gamma^{-2} \int_0^1 \phi^2(u)du\right),$$

where $\gamma = \int_{\mathbb{R}} \varphi(F(x))(-f'(x))dx$.

4.2 Finite sample distribution

Let us firstly introduce some used theorems and definitions.

Definition 11. We say that the function $G(t)$ is called *regularly varying at infinity* with index m, denote $G \in R_m$, if it holds that

$$\lim_{x \to \infty} \frac{G(xt)}{G(x)} = t^m, t > 0.$$

If $m = 0$ the function $G(t)$ is called *slowly varying at infinity*.

Theorem 4.2.1 (Fisher-Tippett)**.** Let $X_1, \ldots, X_n$ be a random sample from nondegerenate distribution function $F(x)$. Then there exists sequences $\{c_n\} > 0$ and $\{d_n\}$ such that for $M_n = \max(X_1, \ldots, X_n)$ it holds

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} Y \tag{4.2.1}$$

where $\mathcal{L}(Y) = H_\gamma$ and

$$H_\gamma = \begin{cases} \Phi_{\frac{1}{\gamma}} & \gamma > 0, \\ \Lambda & \gamma = 0, \\ \Psi_{-\frac{1}{\gamma}} & \gamma < 0. \end{cases}$$

where $\Phi_m(x) = \exp\{-x^{-m}\}, x > 0, m > 0$ is a Frèchet distribution, $\Lambda(x) = \exp\{-e^{-x}\}, x \in \mathbb{R}$ is Gumbel distribution and $\Psi_m(x) = \exp\{-(-x^m)\}, x \leq 0, m > 0$ is Weibull distribution.

Definition 12. We say that the distribution function F is in the *domain of attraction* of H_γ if (4.2.1) holds for some $\{c_n\}$ and $\{d_n\}$. In such case we write $F \in D(H_\gamma)$.

Remark 4.2.1.
1. It can be shown that $\lim_{\gamma\to 0} H_\gamma = H_0$.
2. If $1 - F(x) = x^{-m}L(x)$, where $L(x) \in R_0$, then $F(x) \in D(\Phi_m(x))$.

Lemma 4.2.1 (Von Mises condition)**.** Let $F(x)$ be an absolutely continuous distribution function with density $f(x)$ and let

$$\lim_{x\to\infty} \frac{xf(x)}{1 - F(x)} = m, \quad \gamma > 0,$$

then $F \in D(\Phi_m)$.

Suppose $X_1, \ldots, X_n$ be a random sample from a distribution with symmetric distribution function $F(x - \theta)$. Let F be heavy tailed hence there exists $m > 0$ such that

$$\lim_{x\to\infty} \frac{-\log(1 - F(x))}{m \log x} = 1. \tag{4.2.2}$$

Coefficient m is often called Pareto index in literature. From (4.2.2) we can write

$$\lim_{x\to\infty} \frac{-\log(1 - F(x))}{\log x} = m. \tag{4.2.3}$$

Furthermore it holds that

$$\lim_{x\to\infty} \frac{-\log(1-F(x))}{m\log x} = \lim_{x\to\infty} \frac{\frac{-f(x)}{-(1-F(x))}}{m\frac{1}{x}} = \lim_{x\to\infty} \frac{xf(x)}{(1-F(x))m} = 1,$$

hence

$$\lim_{x\to\infty} \frac{xf(x)}{(1-F(x))} = m.$$

From the von Mises condition it holds that $F \in D(\Phi_m)$, hence we can write that

$$1 - F(x) = x^{-m} L(x),$$

where $L(x)$ is a slowly varying at infinity.

Let T_n be a translation equivariant estimator of θ satisfying assumptions of Theorem 2.2.1. Then we know that

$$\lim_{x\to\infty} \sup B(T_n, a) \leq n.$$

Moreover, suppose

$$\lim_{x\to\infty} B(T_n, a) = \lambda_n. \tag{4.2.4}$$

Now we can deduce the type of distribution T_n. From (4.2.2) and (4.2.4) it holds that

$$\lim_{x\to\infty} \frac{-\log P_0(|T_n| > a)}{\log x} = \lim_{x\to\infty} \frac{-\log P_0(|T_n| > a)}{-\log(1-F(x))} \frac{-\log(1-F(x))}{m\log x} m = m\lambda_n.$$

From uniqueness of limit in (4.2.3) $m\lambda_n$ is a Pareto index of distribution of T_n. Since $m\lambda_n < mn < \infty$, it holds

$$P_0(|T_n| > a) = x^{-m\lambda_n} L_T(x),$$

where $L_T(x)$ is a function slowly varying at infinity. From here it follows that distribution T_n is heavy tailed for every finite n.

Chapter 5

Conclusion

We compared some translation equivariant estimators according to their characteristics of robustness. Sample mean, sample median, trimmed mean, Huber estimator and Hodges Lehmann estimator were concerned. We discussed their breakdown points and described their tail behaviour. All of these estimators except sample mean are robust, however, all of them have heavy tailed distribution when the random sample comes from heavy tailed distribution. Hence we simulated tail behaviour of these estimators when the random sample comes from t-distribution which is heavy tailed, however converges to the normal one for number of degrees of freedom going to infinity, hence converges to the light tailed distribution. These simulation shows that the sample mean and trimmed mean have slowest tails for very small degrees of freedom. The higher the degrees of freedom the estimators with slowest tails become sample mean and sample median. Trimmed mean, Huber and Hodges Lehmann estimator have similarly fast converging tails. However our observation was able to make only for small a since very extreme values of estimators are hard to simulate. For very extreme values of a it still holds that the estimators with very poor tail behaviour are sample mean a trimmed mean.

In this thesis we discussed two types of characteristics of robustness, tail behaviour and breakdown point and their relationship for monotone and translation equivariant estimators. However, these characteristics are not the only one according to which we can describe robustness. Another ones are for example influence function or variance of asymptotic normal distribution of estimators.

Influence function $IF(x; T, P)$ as defined in this thesis can be understand as influence of x on the mistake that we make when we estimate $T(P)$ by $T(P_n)$. If this influence function is unbounded, then the estimator is not obviously robust since x has an infinite influence on mistake of estimator. As far as the variance of asymptotic normal distribution is concerned robust estimators should have low such variance for large family of initial distributions. To investigate the relationship between these other characteristics of robustness could be in the spotlight of future research.

Appendix A

Simulation algorithm - Mathematica

In this Appendix we show algorithms used for simulations in this thesis. We restrict on algorithms for generating values of estimators. Any other information (quantiles, plots) can be easily obtained from data simulated in the following way.

For simulations in this thesis software Mathematica was used because, unlike R, it is able to sample very extreme values from any probability distribution.

1. Firstly, degrees of freedom d, size of the sample n and general vector with elements in count of desired number of generated estimates v are defined.

```
d = 1;
n = 20;
v = Range[10^6];
```

2. Random samples are generated and corresponding estimator is calculated.

a) sample mean

```
Do[v[[i]] = Mean[RandomVariate[StudentTDistribution[d], n]], {i,
   10^6}];
```

b) sample median

```
Do[v[[i]] = Median[RandomVariate[StudentTDistribution[d], n]], {i,
   10^6}];
```

c) trimmed mean with 0.1*n maximum and minimum observations trimmed off

```
Do[v[[i]] =
   TrimmedMean[RandomVariate[StudentTDistribution[d], n], 0.1], {i,
   10^6}];
```

d) Huber estimator

```
huber[data_, iter_] :=
  Module[{min, t, max, newdata, z, size},
   size = Length[data];
   t = Median[data];
   For[m = 1, m <= iter, m = m + 1,
    newdata = data;
    min = t - 1.5; max = t + 1.5;
    For[j = 1, j <= size, j = j + 1,
     If[newdata[[j]] < min, newdata[[j]] = min];
     If[newdata[[j]] > max, newdata[[j]] = max];
     ];
    t = Mean[newdata]
    ];
   t
   ];

For[r = 1, r <= 10^6, r = r + 1,
  s = RandomVariate[StudentTDistribution[d], 100];
  v[[r]] = huber[s, 10]];
```

e) Hodges Lehmann estimator

```
Do[v[[i]] =
   Median[Mean[
     Transpose[
```

```
    Subsets[RandomVariate[StudentTDistribution[d], n], {2}]]]], {i,
  10^6}];
```

3. Function calculating values of $B(T_n, a)$ is defined.

```
vetsi[c_, x_] := If[c > x, 1, 0];
mensi[c_, x_] := If[c < -x, 1, 0];
B[x_] := (Log[(Sum[vetsi[v[[i]], x], {i, Length[v]}] +
       Sum[mensi[v[[i]], x], {i, Length[v]}])/Length[h1b]])/(Log[
  1 - CDF[StudentTDistribution[d], x]])
```

4. Points of $B(T_n, a)$ are calculated for arbitrary maximum of a because of an acceleration of plotting.

```
max=2;
step=0.1;
data = Table[{i, B[i]}, {i, 0, max, step}];
```

List of Symbols

$\mathbb{I}_{[X_i \in A]}$	an indicator function (equals 1 whenever $X_i \in A$ and 0 otherwise)
$\prod_{i=1}^{n} \mathcal{S}$	Cartesian product of sets
$\otimes_{i=1}^{n} \mathcal{B}$	$\sigma(\prod_{i=1}^{n} B_i, B_i \in \mathcal{B})$
$[X \in A]$	$\{\omega \in \Omega : X(\omega) \in A\}$
$[a, b]$	closed interval
$f = O_P(1)$	$\exists$ annular neigh. U of P and $a > 0$ such that $\forall x \in U$ $\|f(x)\| \leq a$
$f = o_P(1)$	$\lim_{x \to P} f(x) = 0$

Bibliography

[1] FIALOVÁ Alena: *Odhad parametru chvostů distribuční funkce.* ROBUST 2000, 42-49, Praha 2001

[2] HE X., JUREČKOVÁ J., KOENKER R, PORTNOY S.: *Tail Behavior of Regression Estimators and Their Breakdown Points.* Econometrica 58, 1195-1214, 1990

[3] JUREČKOVÁ Jana: *Finite-sample Comparison of L-estimators of Location.* Commentationes Mathematicae Universitatis Carolinae 20, 507-518, Prague 1979

[4] JUREČKOVÁ Jana: *Tail-behavior of Location Estimators.* The Annals of Statistics 9, 578-585., Prague 1981

[5] JUREČKOVÁ Jana: *Robustní statistické metody.* Univerzita Karlova v Praze - Nakladatelství Karolinum, 2001,

[6] JUREČKOVÁ Jana, PICEK Jan: *Finite-sample behavior of robust estimators.* WSEAS Press, Venice 2011

[7] JUREČKOVÁ Jana: *Tail-behavior of estimators and of their one-step versions.* Journal de la Société Francaise de Statistique 153, 44-51, 2012

[8] KUŠNIER Jozef, MIZERA Ivan: *Tail Behavior and Breakdown Properties of Equivariant Estimators of Location.* The Annals of the Institute of Statistical Mathematics 53, 244-261., Champaign 2001

[9] ZUO Yijun: *Finite Sample Tail Behavior of Hodges-Lehmann Type Estimator.* Arizona State University, Tempe 2000